

## OPEN ACCESS

## PAPER



## Inferring the drivers of language change using spatial models

## RECEIVED

29 January 2021

## REVISED

7 April 2021

## ACCEPTED FOR PUBLICATION

21 April 2021

## PUBLISHED

20 July 2021

Original content from this work may be used under the terms of the [Creative Commons Attribution 4.0 licence](#).

Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

James Burridge<sup>1,\*</sup> and Tamsin Blaxter<sup>2</sup> <sup>1</sup> School of Mathematics and Physics, University of Portsmouth, Portsmouth PO1 3HE, United Kingdom<sup>2</sup> Gonville and Caius College, Cambridge CB2 1TA, United Kingdom

\* Author to whom any correspondence should be addressed.

E-mail: [james.burridge@port.ac.uk](mailto:james.burridge@port.ac.uk)**Keywords:** statistical physics, spatial models, language evolutionSupplementary material for this article is available [online](#)**Abstract**

Discovering and quantifying the drivers of language change is a major challenge. Hypotheses about causal factors proliferate, but are difficult to rigorously test. Here we ask a simple question: can 20th century changes in English be explained as a consequence of spatial diffusion, or have other processes created bias in favour of certain linguistic forms? Using two of the most comprehensive spatial datasets available, which measure the state of English at the beginning and end of the 20th century, we calibrate a simple spatial model so that, initialised with the early state, it evolves into the later. Our calibrations reveal that while some changes can be explained by diffusion alone, others are clearly the result of substantial asymmetries between variants. We discuss the origins of these asymmetries and, as a by-product, we generate a full spatio-temporal prediction for the spatial evolution of English features over the 20th century, and a prediction of the future.

**1. Modelling language evolution**

Modelling the collective behaviour of systems of sentient agents, from flocks of birds [1, 2], to economies [3], cities [4] and languages [5–7], is attractive, but not easy. To quote Emanuel Derman [8], an early adopter of financial modelling,

*In physics there may one day be a theory of everything; in finance and the social sciences, you're lucky if there is a usable theory of anything.*

The problems which arise when trying to model social systems depend to some extent on the system in question, but share a good deal in common. In many cases we have only one realisation of the change process which we want to understand, and the data available about the change may be sparse. Because we cannot observe the thought processes and motivations of the agents, we caricature them with a simple model. However, many different models may fit observations similarly well. This problem recedes if we have more data and more realisations, which allow us to apply statistical learning methods [9], such as cross validation, to rigorously select the ‘best model’ by maximizing its ability to predict unseen data. If this is not possible, because we have only one realisation of a particular change process, then another approach is to start with some simple default theory, or *null model* [10], and then seek to determine whether it is sufficient to explain observations. We may also consider an *alternative model* with incrementally increased complexity and minimal additional assumptions, the aim being to understand what is missing from the null model, while avoiding unjustified assumptions about what should replace it.

The broad question which interests us is: what are the processes which drive the evolution of languages, and how should they be represented in a coarse grained spatial model? A longer term aim is to be in possession of sets stochastic equations, derived from simple assumptions about speaker-level behaviour, which describe how language evolves over space and time, and amongst different social groups. Such equations should calibrate to available data on the language state up until the present, then allow us to understand historic changes, and make future predictions. Here we take a step toward addressing these aims, by fitting a coarse grained spatial

model to two large scale spatial datasets which capture the states of various English language features at the beginning and end of the 20th century.

The mathematical modelling of language evolution, often using methods of statistical physics, is now an established field. Many aspects of language change have been studied using non-spatial models [5, 11–14], and an important focus has been to understand the replacement of one linguistic feature by another, typically following the ubiquitous temporal S-curve [15, 16]. Such changes can be driven by numerous mechanisms, including those which favour particular variants (e.g. regularization [17], social biases [18], or exogenous factors [16]), by stochastic effects [19, 20], and by biases driven by linguistic variation over the age spectrum [12, 13, 21–23]. In this work we allow speakers a learning bias toward individual linguistic features, without explicitly modelling any particular mechanisms which might generate this. We seek only to determine whether some form of bias existed, given the observed changes.

Spatial modelling of language change is a topic of growing interest [7, 24–31]. Whereas non-spatial linguistic data is plentiful, detailed spatial datasets are rarer. Their construction typically requires either large scale collaborative effort [32–34], or viral success online [35–37]. Spatial modelling presents additional challenges in the form of greater analytical and computational complexity, particularly when matching models to real world spatial language distributions. There is also long standing interest within the quantitative linguistics community in the statistical analysis of spatial, social and temporal variations within real linguistic domains [34, 38, 39].

Recent spatial models [7, 29–31], which accurately model geography and population distributions, suggest that the shapes of linguistic domains and the locations of cities and towns can have a partially predictable influence on the evolution of language. In fact, the importance of geography has been well known to dialectologists for a long time [18, 40]. From a mathematical perspective, this influence arises from the impact of geography on the shapes of boundaries (isoglosses) between alternative linguistic features, which behave like two dimensional bubbles. The existence of this surface tension effect has recently been tested using historical English dialect data [31]. We therefore believe that space and spatial processes must be part of models which seek to understand language change processes. Here we provide, to our knowledge, the first detailed spatial model which matches the evolution of real linguistic features within an extended spatial domain (England in the 20th century), accounting for realistic migration patterns, and plausible learning behaviour.

## 2. Data and existing theories

### 2.1. Linguistic variables and variants

For the purpose of academic study, languages may be broken down into distinct components: single units of sounds (*phones*), rules for combining sounds (*phonology*), words (the *lexicon*), rules for word construction (*morphology*), and rules for sentence construction (*syntax*). A language may be viewed as a complete specification of all its components, and language change as the process of progressively exchanging components for alternatives. We refer to a language feature for which there are a set of alternatives as a *variable*, and to the alternatives as *variants*. These might be different words for the same object or idea, different sounds playing the same role in certain groups of words, or alternative syntactic rules. As an example, the word for the season after summer historically had variants *autumn*, *backend* and *fall* in England [32]. The fundamental quantities which we build models of are the relative usage frequencies of variants at different locations (or regions) in space. Simultaneous spatial variations in many linguistic variables can create distinctive regional *dialects*.

### 2.2. The survey of English dialects and the English dialect app

We model the spatial evolution of a set of linguistic variables over the 20th century in England, the largest and most densely populated nation within the British mainland. As initial and final conditions, we use data for the same variables from two surveys: the survey of English dialects (SED) [41] and the English dialects app (EDA) [35, 36].

The SED was carried out in the 1950s at 313 localities. With 986 respondents, it is a small survey by modern standards. The localities were selected to provide a relatively evenly-spaced sample across England, but included almost no speakers from urban areas. This was a deliberate choice; the survey sought out the most conservative demographic (typically farm-labourers born in the 1870s and 80s) to capture the network of ‘traditional’ dialects before they disappeared. Methodologically, the SED was rigorous, making few allowances for speed or convenience at the expense of data volume and quality: fieldworkers visited respondents in person and recorded their responses to over a thousand questions in narrow phonetic transcription (and later also on tape). Despite some criticisms arising from systematic differences in how fieldworkers asked questions or transcribed responses [42, 43], the SED is a rich and relatively trustworthy source of data and has provided the material for more than half a century’s worth of secondary analysis.

The EDA represents a new generation of dialect data collection methods, using digital technologies to reach a large number of respondents very cheaply. Over 50 000 speakers answered 26 questions about their usage through a smartphone app in 2016, and all but one question duplicated variables surveyed in the SED. The EDA exemplifies a different set of trade-offs between data quality, quantity and speed. It is possible to mine the SED data for patterns across many responses and examine variables which were not specifically targeted, whereas the EDA required speakers to decide between predefined categories, limiting analysis to 26 variables and predetermined sets of variants. Instead, the EDA's major advantage is its 45 287 speakers in 39 590 locations within the region covered by the SED. It is also more representative, covering urban areas and all demographics, although, because it was carried out via smartphones, it is somewhat skewed towards younger and more affluent speakers.

In spite of methodological differences, we argue that comparison of the SED and EDA is a viable way to explore language change in 20th century English. Laboratory perception tests [36] suggest that speakers are largely able to discriminate between the variants of the phonetic variables elicited in the EDA. Spatial patterns in the results of the EDA, when compared to the SED, show that levelling (loss of variants) and isogloss (linguistic boundary) movements are in line with our prior understanding of the mechanisms of language change. However, because typical SED respondents were old, while EDA respondents were younger, we are not simply mapping the 60 or so years of language change between the median dates of the two language surveys. If we assume that speakers' linguistic norms are set early in life and change little in adulthood (the underlying assumption of much historical linguistic [44] and sociolinguistic [45] work on language change), the relevant comparison is instead speakers' dates of birth: the median date of birth is the 3rd of May 1881 for SED respondents, and the 20th of July 1983 for EDA respondents, implying that we are looking at around 100 years of change.

Because urban areas are under-represented in the SED, we must consider whether city varieties are likely to have differed substantially from the varieties in the regions surrounding them. In that case SED maps would be missing islands of highly divergent usage in places where spatial models [7, 29, 30] suggest that population density gradients should preserve those distinctions or encourage the spread of urban variants. However, the rapid urbanisation of the industrial revolution had started only a couple of generations before the SED speakers were born; less than the three generations required for new dialect formation, according to standard models [46], and certainly not long enough for urban varieties to have substantially diverged from their regional inputs. This is borne out by evidence. For example, Coates [47], surveying research on traditional Bristol English, notes: 'what can be shown to distinguish Bristolian at all linguistic levels from other dialects of the region is relatively little'. Clearly city varieties at the time of the SED will have had distinguishing features, just as varieties at any point in space differ in some ways from those nearby; but we can assume that they tended to agree in most respects, and so their omission is not likely to be any more problematic than a situation where any other location happened to be lacking in samples.

### 2.3. Current hypotheses about linguistic changes

The overwhelming story of change in English dialects over the 20th and early 21st centuries is one of loss of diversity. In the words of Britain [48]:

*There has been such considerable and ongoing dialect attrition that the language use reported across the country by Ellis's survey of 1889 seems, in many cases and in many places, quite alien to that spoken just over 100 years later.*

In place of smaller traditional dialects we find distinctions at larger, regional levels [49–51], and much research has focused on these *regiolects* and the processes of regional dialect levelling [49, 52–55] which generated them. Reductions in geographical linguistic differences are not only due to convergence to one of several local variants but also to geographically widespread adoption of common innovations [50]. These changes are part of a near-universal pattern across the traditional varieties of Europe [56–58]. It is believed that interactions between individuals can lead to a decline in linguistic variation via *accommodation*, where conversation partners adjust their speech to better match each other, and by child learners acquiring accommodated forms [59–63]. However, in certain social contexts, children may also learn variants directly from mobile outsiders [64]. The fact that accommodation is mediated through interactions between speakers has led some linguists to conclude that the primary drivers in the decline of linguistic diversity are travel, commuting and migration [49, 50, 55, 59, 65–67]. Other potential drivers discussed include changes in social network structure [49, 50, 55, 59, 65, 66, 68], the age structure of the community [69], the influence of mass media [58, 70], normative attitudes and education [58, 71, 72] and relatedly the salience and stereotype status of particular variants [55, 73], identity factors [74], the informalisation of public life [58] and socio-economic forces [55]. Purely linguistic *internal* factors such as structural regularity, functional economy, or naturalness may determine which variants win out in the levelling process [49, 55, 59, 67] or they may not be relevant at all [50, 71].

Arguments have also been made for the central importance of ideological factors, in particular strong normative attitudes towards the standard, alongside mobility and contact [72, 74, 75]; parallel to these arguments, it has been suggested that strong alignment of speaker identities with the local community may be enough to check the levelling process and so preserve a distinct local variety [51, 58].

There are clearly numerous potential drivers of linguistic change, and it is beyond the scope of this work to quantify their relative importance. It may be that driving processes which operate in different ways at small scales (at the level of motivations, interactions and contacts between individuals) yield similar or identical terms in evolution equations for coarse grained population averages, making the task of inferring the importance of individual effects impossible using such models. However, macroscopic models have been used to infer different *classes* of driving mechanism from non-spatial linguistic time series, specifically exogenous (population level) vs endogenous (individual level) drivers [16]. We take a similar approach here, but our two classes are spatial processes (migration, movement), and processes which introduce asymmetry with respect to variants (ideology, internal linguistic effects, social prestige, normative bias toward a standard etc). We do this by defining coarse grained evolution equations which account for both movement and biased copying. We then fit our model to the initial and final conditions supplied by the SED and EDA data, to determine the relative importance of each process. Since movement is often viewed as the primary driver of change, we view variant symmetric dynamics as our null model, and determine the extent to which it can explain the observed changes, before breaking variant symmetry by allowing for biasing factors.

### 3. The model

We consider a spatial domain divided into  $L$  cells, each containing (approximately)  $N$  speakers. The centroids of these cells are written  $\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_L$ . Consider a linguistic variable with  $q \in \{1, 2, \dots\}$  variants, and let the relative frequencies with which these variants are used within cell  $\mathbf{r}$  define the frequency vector  $\mathbf{f}(\mathbf{r}) = (f_1(\mathbf{r}), \dots, f_q(\mathbf{r}))^T \in \Delta^q$ , where  $\Delta^q$  is the  $q$ -dimensional simplex and  $^T$  denotes transpose. We have suppressed time dependence for brevity. The cells we use in our analysis are middle layer super output areas (MSOAs), a set of  $L = 7,201$  geographical polygons, each with a similar number of residents, used for reporting census data in England and Wales [76]. The mean MSAO population in 2011 was 7,787, and the modal area of English MSOAs is approximately  $0.29 \text{ km}^2$ , corresponding to a few tens of streets within a densely populated city. Interactions with areas outside the English borders, for which we lack survey data, are neglected. We justify this simplification on the basis that population densities are low at the Scottish and Welsh borders, the English make up the great majority ( $\approx 87\%$ ) of the British mainland population, and Wales and Scotland possess distinctive cultural and linguistic identities. We assume that much longer range interactions, for example from American TV, cinema, and music, will be subsumed into biases in the *learning function* (see section 3.3) which determines the adoption probabilities of different variants by young speakers.

#### 3.1. The language community

We refer to the language that a speaker is exposed to as their *linguistic environment*, and assume this environment consists predominantly of voices from their own and nearby cells. We allow for biasing factors by weighting the influence of variants using a set of biases which affect the learning process (see section 3.3). We capture the local environment of a speaker in cell  $\mathbf{r}$  using the *community influence matrix*,  $W$ , where  $W(\mathbf{r}, \mathbf{r}')$  gives the influence that speakers from  $\mathbf{r}'$  have on speakers from  $\mathbf{r}$ . The matrix is *stochastic*, meaning that it is square with unit row sums. We define the *community frequency vector* of a variable as follows

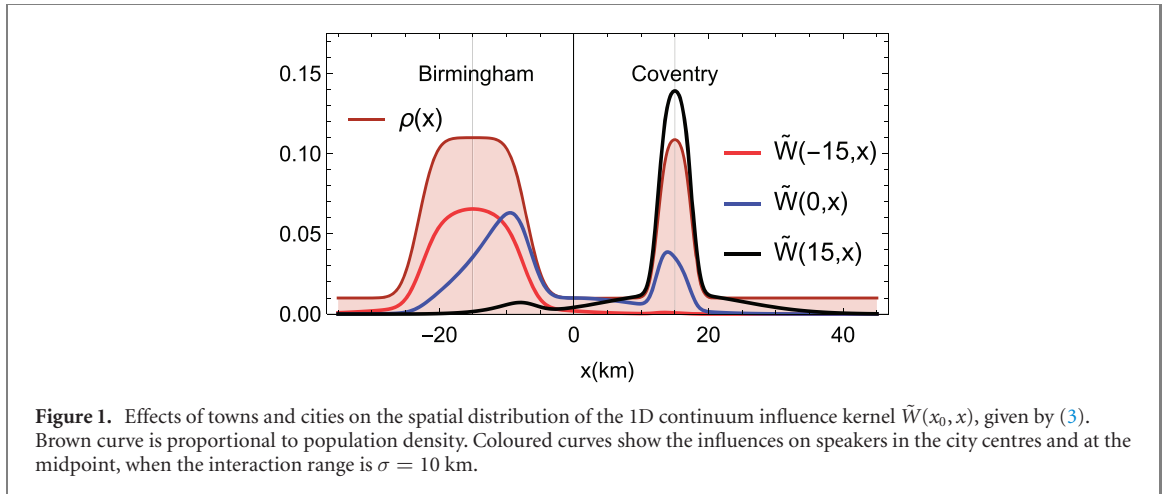
$$\hat{\mathbf{f}}(\mathbf{r}) \triangleq \sum_{\mathbf{r}'} W(\mathbf{r}, \mathbf{r}') \mathbf{f}(\mathbf{r}'), \quad (1)$$

where  $\triangleq$  denotes a definition. We assume that local influence strengths are mediated by two factors: physical separation, and population distribution. A definition which incorporates both is

$$W(\mathbf{r}, \mathbf{r}') = \frac{\exp\left(-\frac{|\mathbf{r}-\mathbf{r}'|^2}{2\sigma^2}\right)}{\sum_{\mathbf{r}''} \exp\left(-\frac{|\mathbf{r}-\mathbf{r}''|^2}{2\sigma^2}\right)}, \quad (2)$$

where we call  $\sigma$  the *interaction range*. To illustrate the competing effects of interaction range and population distribution we consider the one dimensional small-cell limit of the influence matrix, which takes the form of a kernel giving the influence of location  $x$  on location  $x_0$

$$\tilde{W}(x_0, x) = \frac{\rho(x) \exp\left(-\frac{(x-x_0)^2}{2\sigma^2}\right)}{\int \rho(y) \exp\left(-\frac{(y-x_0)^2}{2\sigma^2}\right) dy}, \quad (3)$$



where  $\rho(x)$  is population density at  $x$ , inversely proportional to cell size at this position. Population density is automatically accounted for in definition (2) because higher density areas have more cells. Figure 1 shows the spatial distribution of influences on three speakers who live in or between two cities with radii and separation similar to Birmingham (population  $\approx 10^6$ ) and Coventry (population  $\approx 3 \times 10^5$ ). Consider a speaker in central Coventry ( $x_0 = 15$ ). Most of the influence on them comes from within the city, because there are relatively few people to interact with outside. This makes their effective interaction range smaller. A speaker in central Birmingham ( $x_0 = -15$ ) will have a wider range of influence because their city is larger. For a speaker in between the two cities ( $x_0 = 0$ ), influence comes from both, with the larger city dominating (see blue curve in figure 1). In the limit of very large cities (or other regions of approximately uniform population density) the influence distribution becomes normal so that  $\approx 63\%$  of influence comes from within  $\sigma$  km.

We assume that speakers' linguistic states are fixed in their youth (see section 3.3), so the interaction range should be chosen to capture the community state observed by a younger speaker. An empirical measure of such separations is provided by school travel distances; secondary education has been compulsory since 1918, so we would expect children to regularly interact with others from their school catchment. In the early 21st century secondary pupils travelled, on average,  $\approx 5.5$  km to school [77]. We will call this the *mean catchment radius*, denoted  $\mu_R$ . The number of state schools has declined somewhat over the 20th century; in 1951 there were 5900 vs 4072 in 2010 [78]. This implies that average catchment radii have increased, with 1951 radii being  $\approx 83\%$  early 20th century radii. Scaling 2010 travel distances by this factor yields an average travel distance of  $\approx 4.5$  km. Suppose we model the distribution of displacements between home and school as a bivariate normal distribution

$$f(x, y) = \frac{e^{-\frac{x^2 + y^2}{2\omega^2}}}{2\pi\omega^2}, \quad (4)$$

then the mean home-school distance is  $\mu_R = \sqrt{\frac{\pi}{2}}\omega$  and the mean distance between two homes is  $2\omega = \sqrt{\frac{8}{\pi}}\mu_R \approx 1.6\mu_R$ . This suggests a typical distance between interacting speakers is between 5 and 10 km. Since  $\sigma$  functions as an upper bound in cities, we take our interaction range as the upper limit of this interval, so  $\sigma = 10$  km  $\approx 6$  miles.

We assume that speakers from a given cell have similar environments, so that  $\hat{\mathbf{f}}(\mathbf{r})$  approximates the perceptions of all speakers in zone  $\mathbf{r}$ . The assumption that all the speakers in a cell access the same primary linguistic data is a simplification of reality. We know that within many geographical areas there are subgroups whose language differences may depend on class, sex or ethnicity [18]. We simplify in this way because we are interested in national level spatial variations for which we have detailed geographical data, but no consistent breakdown by social subgroup, so cannot infer any effects of subgroups. We implicitly assume that social subgroups can induce biases on variants.

### 3.2. Migration

Migratory processes may be *internal*, where individuals move between locations in a single language area (typically a country), or *external*, where they arrive from another country or language area. During the 20th century, foreign-born people have made up a small and slowly changing percentage of the English population; 4.2% in 1951, for example, rising to 6.7% in 1991 [79]. The cumulative total number of migrants to Britain from the early 1800s to 1945 is estimated as 2.34 million [80]. Since numbers are small, we do not



explicitly model external migration, instead viewing it as one of many possible factors which may induce biases. To model internal migration, we need to know how often people move, and how far. The English Longitudinal Survey of Ageing (2007), showed that for individuals aged 50–89, the average number of different residences occupied for at least six months during their lifetime was 5.6 for those born 1918–1927, rising to 6.42 for those born 1948–1957 [81], giving an average of approximately one move every ten years. Inter-county migration distances may be extracted from birthplace and place of residence recorded in censuses, providing data back to the mid nineteenth century [82]. Such data is, however, spatially coarse; it provides no information about short range migrations. Modern lifestyle survey data can help fill the gap. Analysis [83] of large scale research polls by Acxion Ltd (2005–2007), which recorded current and previous postcode, generated  $\approx 1.25 \times 10^5$  migration distances, having mean 25.77 km, median 2.89 km and standard deviation 63.91 km. This implies a migration distribution concentrated at short distances, with sub-exponential decline at large distances. This large distance migration behaviour is often modelled with a power, or ‘gravity’ law [83, 84].

Models of migration must be consistent with migration distance statistics, and constrained so that total flows in and out of cells match observed values. This can be achieved [85] by writing the expected number of people who leave their address in cell  $\mathbf{r}$  and move to a new address in  $\mathbf{r}'$  as

$$m(\mathbf{r}, \mathbf{r}') \triangleq A(\mathbf{r})B(\mathbf{r}')O(\mathbf{r})D(\mathbf{r}')h(|\mathbf{r} - \mathbf{r}'|), \quad (5)$$

where  $O(\mathbf{r})$  is the total number of movers whose previous address was in  $\mathbf{r}$ ,  $D(\mathbf{r}')$  is the total number of people whose new address is in  $\mathbf{r}'$ , and  $h(|\mathbf{r} - \mathbf{r}'|)$  is a distance function which, typically, decreases for longer moves. The conditions

$$\sum_{\mathbf{r}'} m(\mathbf{r}, \mathbf{r}') = O(\mathbf{r}) \quad (6)$$

$$\sum_{\mathbf{r}} m(\mathbf{r}, \mathbf{r}') = D(\mathbf{r}'), \quad (7)$$

imply that

$$A(\mathbf{r}) = \left( \sum_{\mathbf{r}'} B(\mathbf{r}')D(\mathbf{r}')h(|\mathbf{r} - \mathbf{r}'|) \right)^{-1} \quad (8)$$

$$B(\mathbf{r}') = \left( \sum_{\mathbf{r}} A(\mathbf{r})O(\mathbf{r})h(|\mathbf{r} - \mathbf{r}'|) \right)^{-1}. \quad (9)$$

A mover might shift to a new address in the same cell, and the quantity  $m(\mathbf{r}, \mathbf{r})$  gives the expected number of people who do this within cell  $\mathbf{r}$ . The sets of constants  $A \triangleq \{A(\mathbf{r}_k)\}_{k=1}^L$  and  $B \triangleq \{B(\mathbf{r}_k)\}_{k=1}^L$  are found by generating a sequence of sets  $A_0, B_0, A_1, B_1, \dots$  where  $A_0$  is the set for which  $A(\mathbf{r}) = 1$  for all  $\mathbf{r}$ ,  $B_k$  is computed from  $A_k$  using (9), and  $A_{k+1}$  is computed from  $B_k$  using (8). The sequence is then iterated to convergence.

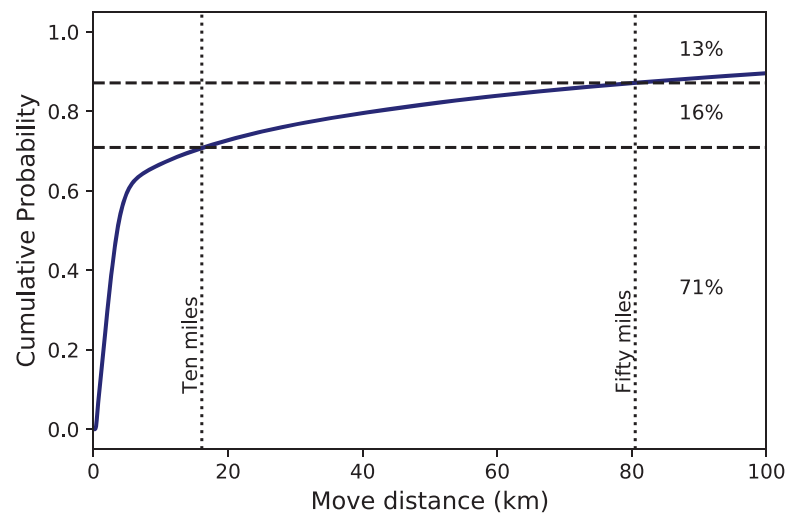
We assume that our cell populations are in equilibrium and without loss of generality assume there is one mover per cell, so  $O(\mathbf{r}) = D(\mathbf{r}) = 1$ , and  $m(\mathbf{r}, \mathbf{r}')$  is a stochastic matrix with elements which represent transition probabilities for a single mover. The matrix can be rescaled to capture realistic migration rates. We define the distance function

$$h(r) = \omega e^{-(r/a)^2} + \mathbf{1}_{\{r>0\}}(1 - \omega)r^{-\rho}, \quad (10)$$

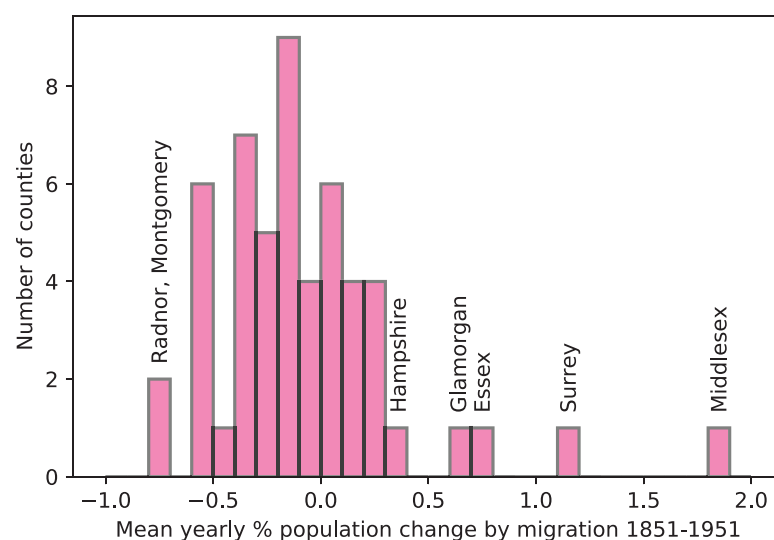
where the first term models short range moves with typical range  $a$ , and the second term is the long distance fat-tail of the distribution, with gravity exponent  $\rho$ . The parameter  $\omega \in [0, 1]$  interpolates between entirely short range, and pure power law distributions. Having calculated  $m(\mathbf{r}, \mathbf{r}')$  for our system, the cumulative distribution of the migration distance  $X$  is calculated as

$$\mathbb{P}(X < r) = \frac{1}{L} \left[ \sum_{\substack{\mathbf{r}, \mathbf{r}' \\ \mathbf{r} \neq \mathbf{r}'}} H(r - |\mathbf{r} - \mathbf{r}'|) m(\mathbf{r}, \mathbf{r}') + \sum_{\mathbf{r}} H\left(r - R(\mathbf{r})/\sqrt{2}\right) m(\mathbf{r}, \mathbf{r}) \right], \quad (11)$$

where  $H$  is the Heaviside step function and  $R(\mathbf{r})$  is the radius of a circle with identical area to cell  $\mathbf{r}$ . The second term in the square brackets accounts for all intra-cell moves, assumed to be of distance  $R(\mathbf{r})/\sqrt{2}$ , following [86]. We select distance function parameters to match the short range (median) and long range (standard deviation) statistics of the inter- and intra-cell distance distribution computed in [83] for MSOAs, using the Acxion data. This was achieved by parameter grid search to minimize the sum of the relative errors



**Figure 2.** Cumulative distribution of migration distances generated using distance function (10) with  $\omega = 0.75$ ,  $a = 3$  km and  $\rho = 1.5$ . The mean, median and standard deviation are 31.3 km, 3.6 km and 63.4 km respectively. For comparison, the corresponding statistics computed from Acxion survey data [83] are 26.3 km, 3.6 km, and 63.7 km. Horizontal dashed lines show the fraction of moves less than ten miles (71%) and fifty miles (87%).



**Figure 3.** Yearly percentage changes in population for English and Welsh counties, averaged over 1851–1951, calculated from the census based estimates of Friedlander [82]. Counties with the highest and lowest average rates of change are labelled.

in the median and standard deviation between the model and the Acxion data. The resulting distribution (figure 2) is our estimate for migration distances at the turn of the 21st century. Its shape reflects two kinds of migratory events: frequent short moves, perhaps to change accommodation, and rarer long distance moves, leaving behind friends, work and other social contacts, where the distance travelled becomes relatively less significant.

We now consider whether this distribution may be used to approximate migration during in earlier periods. Increased move frequency over the 20th century does not imply that the distribution of distances has also changed. Although we lack detailed information about historical short range moves, we can compare our distribution to statistics of longer migrations extracted [87] from the National Marriage and Fertility survey (1959–1960), which sampled more than 2000 couples aged 16–60. The survey showed that 16% of moves were of 10–49 miles and 15% were 50 miles or more. Our corresponding percentages are 16% and 13% respectively (figure 3). The similarity of these estimates, and the sample error associated with historical percentages (estimated at  $\pm 1\%$ ), suggests that the statistics of long distance internal migrations have not changed dramatically. Because people are most likely to migrate in their mid twenties [81], many of the moves in the 1959–1960 marriage survey may have been made much earlier in the century.

We also consider whether our assumption that each cell is in equilibrium is reasonable. Using estimates of net inter-census population changes calculated by Friedlander [82] for counties in England and Wales, we have computed the yearly percentage population change by county, averaged over 1851–1951 (figure 3). The results show south east England experienced the greatest inflow of migrants, and that the root mean squared net flow rate over all counties is  $\approx 0.45\%$ . By comparing this figure with the overall yearly migration rate of  $\approx 10\%$  we see that, on average, in cells that are gaining, for every 20 people who leave, 21 will arrive. Since net changes in population are therefore typically at least an order of magnitude smaller than absolute population flows, we assume that non-equilibrium effects can be ignored.

Long range migration in principle allows variants to jump between locations within the spatial domain. This jumping process is the subject of the linguistic *gravity model* [18, 88], which attempts to quantify the spread of variants between a finite set of population centres, viewed as points in space. In contrast, the cells in our model cover the entire spatial domain, and population centres appear as densely packed clusters. Other than cell size, there is no intrinsic difference between city and rural cells. The effects of allowing city speakers to behave differently to their rural counterparts, in a spatial model similar to ours, was explored in [29], along with connections to other classical models of spatial linguistic spread including the wave model and hierarchical diffusion [40, 89].

### 3.3. Language learning and evolution

We assume an iterative model of language evolution, in which dying adults are replaced by new speakers who select a variant using the probability mass function

$$\mathbf{p}(\mathbf{r}) \triangleq \mathbf{g}(\hat{\mathbf{f}}(\mathbf{r})), \quad (12)$$

where  $\mathbf{g}: \Delta^q \rightarrow \Delta^q$  is the *learning function*. An alternative interpretation, which yields dynamics which have identical deterministic component, but no stochasticity, is that the components of  $\mathbf{p}(\mathbf{r})$  give the relative frequencies with which new speakers use different variants. We could, if we wished, interpolate between these two interpretations by assuming a mixture of the two behaviours. In either case equation (12) captures the learning process which converts infants into adult speakers. It is an approximation of the cumulative effect of countless small iterations made within a changing learning environment. Underlying this model is the assumption that linguistic behaviour is mainly acquired in childhood, and changes little in later life. Learning to speak requires the assimilation of a variety of structures, from sets of individual sounds and sound patterns, through words and morphological rules, to syntax. Learning processes differ between structures, and we do not aim to capture these processes in detail. Rather we propose a simple but plausible learning function which maps current speech patterns to learned forms. We derive evolution equations by considering the case which generates maximum stochasticity: where new speakers select a single variant. We will then consider the importance of this stochasticity.

A simple learning model, analogous to *neutral evolution* in genetics [5, 14, 90–92], is that variants are selected with a probability equal to their current frequency. This has two problems as a model of language evolution. First, as we will show below, linguistic communities would take an extraordinarily long time to settle on one feature. Second, the geographical boundaries between language features (isoglosses) revealed by linguistic surveys, appear to require some form of social *conformity* [7, 29, 31] meaning that speakers preferentially select the most common variants with a probability which *exceeds* their relative frequency. Such a selection strategy is optimal if advantage can be gained by matching the speech of others. We can incorporate both conformity and asymmetric bias toward different variants by defining the learning function

$$[\mathbf{g}(\mathbf{f})]_k \triangleq \frac{h_k f_k^\beta}{\sum_{i=1}^q h_i f_i^\beta}, \quad (13)$$

where  $[\mathbf{f}]_k$  denotes the  $k$ th component of the vector  $\mathbf{f} \in \Delta^q$ . The variables  $h_k > 0$  are biases, and  $\beta \geq 1$  is the *conformity number*. If  $\beta > 1$  then  $\mathbf{g}(\mathbf{f})$  increases the frequencies of already popular variants, and reduces the frequencies of less popular ones. The biases  $h_1, \dots, h_q$ , which we write as a vector  $\mathbf{h}$ , introduce variant asymmetry, with the selection probability of the  $k$ th variant an increasing function of  $h_k$ . However,  $\mathbf{g}(\mathbf{f})$  is invariant under a constant re-scaling of all biases so it is their ratios which determine the extent to which variants are favoured.

To derive equations for the evolution of frequencies, we denote the counts of speakers using the variants of a given linguistic variable within cell  $\mathbf{r}$  at discrete time step  $t$  as

$$\mathbf{X}(\mathbf{r}, t) = (X_1(\mathbf{r}, t), \dots, X_q(\mathbf{r}, t))^T, \quad (14)$$

where



$$\sum_{i=1}^q X_i(\mathbf{r}, t) = N. \quad (15)$$

We model *renewal* (birth and death), and *migration*. At each time step,  $n_R$  adults die and are replaced with young speakers, and  $n_M$  migrate. The renewal and migration rates are

$$\lambda_R \triangleq \frac{n_R}{N} \quad (16)$$

$$\lambda_M \triangleq \frac{n_M}{N}. \quad (17)$$

We set  $\lambda_M = 0.1$ , corresponding to one migration per decade, and  $\lambda_R = 0.02$  corresponding to a typical period of 50 years between reaching linguistic maturity, and leaving the language community. We justify this latter estimate on the basis that linguistic maturity occurs in the late teens [93], and that UK life expectancy in the mid 20th century was in the late sixties [94]. The total number of speakers who die or migrate in each cell is  $n \triangleq n_R + n_M$ . These speakers are selected uniformly at random from the existing population, implying that both lifetimes and times between migrations are geometrically distributed with parameters  $\lambda_R$  and  $\lambda_M$ , respectively.

If each migrator were to select their new cell using the symmetric transition probability matrix  $m(\mathbf{r}, \mathbf{r}')$ , then the expected number of migrators received by each cell would be  $n_M$ , and the expectation of the mean linguistic state of the migrators incident on cell  $\mathbf{r}$  would be

$$\sum_{\mathbf{r}'} m(\mathbf{r}', \mathbf{r}) \mathbf{f}(\mathbf{r}', t) \triangleq \bar{\mathbf{f}}(\mathbf{r}, t). \quad (18)$$

We model migration by replacing our  $n_M$  migrators with  $n_M$  speakers with states selected according to the probability vector  $\bar{\mathbf{f}}(\mathbf{r}, t)$ . We write the variant counts of these speakers  $\mathbf{X}_M(\mathbf{r}, t) \sim \text{multinomial}(n_M, \bar{\mathbf{f}}(\mathbf{r}, t))$ . The variant counts of the speakers who replace the dead is a random variable,  $\mathbf{X}_R(\mathbf{r}, t) \sim \text{multinomial}(n_R, \mathbf{p}(\mathbf{r}))$ . The variant counts of the speakers who die or migrate is a random variable,  $\mathbf{Y}(\mathbf{r}, t)$ , drawn from the multi-hypergeometric distribution [95] with parameters  $(n, \mathbf{X}(\mathbf{r}, t))$ . The change in variant counts between  $t$  and  $t + 1$  is then

$$\Delta \mathbf{X}(\mathbf{r}, t) = \mathbf{X}_R(\mathbf{r}, t) + \mathbf{X}_M(\mathbf{r}, t) - \mathbf{Y}(\mathbf{r}, t), \quad (19)$$

where  $\Delta \mathbf{X}(\mathbf{r}, t) \triangleq \mathbf{X}(\mathbf{r}, t + 1) - \mathbf{X}(\mathbf{r}, t)$ . Dividing by  $N$ , we have

$$\Delta \mathbf{f}(\mathbf{r}, t) = \mathbb{E}[\Delta \mathbf{f}(\mathbf{r}, t)] + (\Delta \mathbf{f}(\mathbf{r}, t) - \mathbb{E}[\Delta \mathbf{f}(\mathbf{r}, t)]) \quad (20)$$

$$= \lambda_R \left( \mathbf{g}(\bar{\mathbf{f}}(\mathbf{r}, t)) - \mathbf{f}(\mathbf{r}, t) \right) + \lambda_M (\bar{\mathbf{f}}(\mathbf{r}, t) - \mathbf{f}(\mathbf{r}, t)) + \Delta \mathbf{Z}(\mathbf{r}), \quad (21)$$

where  $\Delta \mathbf{Z}(\mathbf{r}) \triangleq \Delta \mathbf{f}(\mathbf{r}, t) - \mathbb{E}[\Delta \mathbf{f}(\mathbf{r}, t)]$  is a zero mean noise term. The variance of the  $k$ th component of this term, using the properties of the multinomial and multi-hypergeometric distributions, and suppressing  $\mathbf{r}$  dependence for brevity, is

$$\mathbb{V}[\Delta Z_k] = \frac{1}{N} \left( \lambda_R p_k (1 - p_k) + \lambda_M f_k (1 - f_k) + \frac{\lambda (1 - \lambda)}{1 - N^{-1}} f_k (1 - f_k) \right), \quad (22)$$

where  $\lambda \triangleq \lambda_R + \lambda_M$ . The magnitude of the stochastic element of cell dynamics therefore scales as  $N^{-1/2}$ . For reasons set out below, we assume  $\Delta \mathbf{Z}(\mathbf{r}) \approx 0$ , in the current paper.

### 3.4. The importance of noise

To explore the significance of the noise term in (21), we consider a small, well connected community where linguistic evolution is neutral [90, 91, 96], meaning that new speakers adopt variants with probabilities equal to their relative frequencies, so  $\mathbf{p}(\mathbf{r}) = \mathbf{f}(\mathbf{r})$ . In the case of a binary variable, the relative frequency of variant 1 admits the diffusion approximation [97]

$$df = \sqrt{\frac{2\lambda_R}{N}} \sqrt{f(1-f)} dW_t, \quad (23)$$

where  $W_t$  is a standard Brownian motion [98]. Equation (23) is known as the Wright–Fisher diffusion [97] and describes the evolution of the relative frequency of one of two variants (alleles) of a gene in a haploid organism. Notice that the evolution is driven entirely by noise in this case. Let  $T$  be the time taken for the cell

population to settle on one variant ( $f = 0$  or  $1$ ). This is the *fixation time*. Starting from the initial condition  $f = \frac{1}{2}$ , we have

$$\mathbb{E}[T] = \frac{N}{\lambda_R} \ln 2. \quad (24)$$

To be concrete, let us consider the case where  $N$  is the average population of a single cell ( $N = 7,787$ ) in our model, equivalent to a  $0.25 \text{ km}^2$  block of city streets. Using  $\lambda_R = 0.02$  (as above) we have  $\mathbb{E}[T] = 269,877$  years. From this we see that even in a relatively tiny population, the time for one variant to become dominant by noise alone is comparable to the time for which humans have existed as a species. We therefore neglect the stochastic element of our dynamics on the basis that it is inadequate to explain changes over a plausible time scale. Many of the linguistic variants which interest us have disappeared over the course of a single century and only the deterministic terms in (21) are capable of producing such a change.

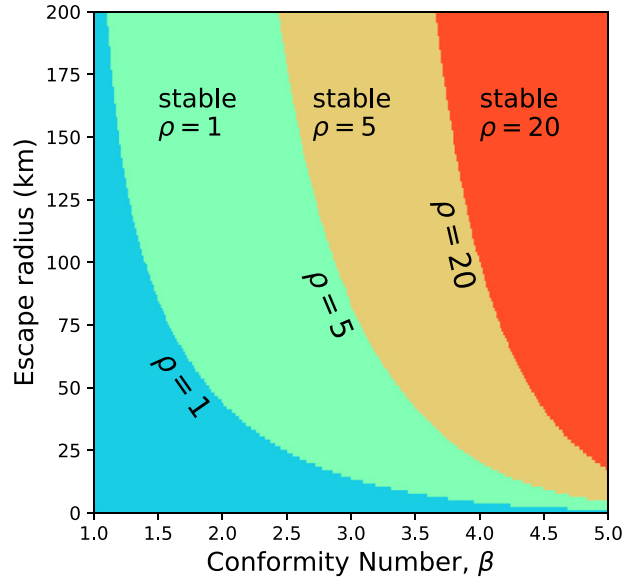
Having decided to neglect the stochastic term in our model, we now consider whether other forms of stochasticity could explain the observed changes. We have assumed our small community is ‘well connected’, meaning that all learners are exposed to the same inputs. Suppose instead that children learn variants from only a small subset of the community. This effectively subdivides the population into smaller groups (e.g. families) within which variants are reproduced by neutral selection, and between which individuals are exchanged (e.g. by marriage). Such processes are studied in models of neutral genetic evolution in subdivided populations, such as Wright’s island model [99], and Kimura’s stepping stone model [91, 96]. At low exchange rates between groups, the time to fixation increases relative to the fully mixed (single group) case. At sufficiently high mixing rate the population becomes effectively panmictic, and fixation time statistics match the fully mixed case. Subdividing the population therefore does not change our decision to neglect the stochastic term in our model. One way for unbiased random copying to be a more powerful driver of change is via the topology and dynamics of the social contact network [92]. For example, if it is strongly clustered then certain individuals or groups become particularly influential. Another possibility is that variants spontaneously gain linguistic *momentum* [12, 21, 100] starting from small stochastic fluctuations. We do not rule out either possibility, but emphasize that stochasticity as we have modelled it is inconsequential in large populations. This does not mean that we think language change is deterministic. Rather, unpredictable driving forces must result from factors which are not part of our model. Whether their origin is stochastic momentum-like effects, social network dynamics, internal linguistic biases or societal changes outside the linguistic system, they all exert an effective bias (real or apparent) on language change, but in a direction that cannot be known in advance. We can view these factors as random variables whose values are realised as history unfolds. Having observed our linguistic system’s historical state, we can then ask: how strong must these biases have been, acting in combination with migration, mixing and social conformity, to produce today’s language distributions? Having established which variants require bias in our simple model, we can then turn to understanding their origin.

### 3.5. Interpreting model parameters

Our deterministic evolution equation

$$\Delta \mathbf{f}(\mathbf{r}, t) = \lambda_R \left( \mathbf{g}(\hat{\mathbf{f}}(\mathbf{r}, t)) - \mathbf{f}(\mathbf{r}, t) \right) + \lambda_M (\bar{\mathbf{f}}(\mathbf{r}, t) - \mathbf{f}(\mathbf{r}, t)) \quad (25)$$

models four processes, two spatial, two not. The two right-hand terms are the *replacement term*, with rate constant  $\lambda_R$ , and the *migration term* having rate  $\lambda_M$ . Mundane mobility is represented by community matrix  $W(\mathbf{r}, \mathbf{r}')$  in our model, which is used to compute  $\bar{\mathbf{f}}(\mathbf{r}, t)$  and migration by the matrix  $m(\mathbf{r}, \mathbf{r}')$ , used to compute  $\hat{\mathbf{f}}(\mathbf{r}, t)$ ; the effects of these are modulated by conformity  $\beta$ . Together, these represent a purely spatial, accommodation- and diffusion-driven model which does not differentiate between variants. Factors which might create bias (see section 2.3) are modelled with the bias vector  $\mathbf{h}$ . If dialect change has mostly been driven by mobility, as many scholars have suggested, we would expect  $\mathbf{h} \approx \mathbf{1}$ . However, if linguistic factors have been crucial in selecting variants, if the standard ideology has driven levelling, or if local identity factors are important in explaining differences between conservative and innovative regions, then  $\mathbf{h}$  will play a more substantial role. One possibility is that bias drives the selection of the levelling target, but that migration and other spatial factors determine the speed and course of levelling. This would be reflected in the model by high but spatially uniform  $\mathbf{h}$ . In such a case, bias might reflect a linguistic constraint derived from a property common to the linguistic system at all localities, or a near-universal social pressure (such as normative pressure from universal schooling). Alternatively, observed regional differences in levelling in the EDA data might not be explained by migration and spatial factors alone and must be modelled with a strongly spatially varying bias. This would imply that local differences in language ideologies play an important role, or that interactions between dialect features create internal pressures varying between dialects.



**Figure 4.** Phase diagram showing regions of  $(\beta, r)$  space for which  $\mathcal{D}$  can retain variant 1 when  $\phi_1 = 0$ , for migration–replacement ratios,  $\rho = \lambda_M/\lambda_R$ , in the set  $\rho \in \{1, 5, 20\}$ . If a region of the phase diagram is stable for some  $\rho^*$ , then it is stable for all  $\rho < \rho^*$ .

#### 4. Interaction between conformity and migration

In the absence of bias, when  $\beta > 1$ , then the replacement and migration terms in (25) represent two competing processes. The replacement term moves the community toward a single local variant, whereas the migration term pulls local forms toward the average linguistic state in a much wider linguistic area. Before comparing our model to data, we illustrate how it operates.

##### 4.1. Survival of an isolated domain

We consider a binary variable, and let  $\mathbf{f}(t) = (f_1(t), f_2(t))^T$  be its relative frequency vector within a dialect area  $\mathcal{D}$ . Let  $r$  be the *escape distance*: the expectation of the distance from a randomly chosen point in  $\mathcal{D}$ , to the boundary of  $\mathcal{D}$ , in a randomly selected direction. The fraction of migrators who escape  $\mathcal{D}$  is then approximately  $1 - \mathbb{P}(X < r) \triangleq \epsilon(r)$  where  $\mathbb{P}$  is the migration measure defined in (11). By the symmetry of the migration matrix, this is equal to the fraction of speakers who enter  $\mathcal{D}$ . Writing  $\phi_1$  for the relative frequency of variant 1 outside  $\mathcal{D}$ , then the average state of the migrators incident on  $\mathcal{D}$  (including internal movements) will be

$$\bar{f}_1 = \underbrace{(1 - \epsilon(r))f_1}_{\text{moves within } \mathcal{D}} + \underbrace{\epsilon(r)\phi_1}_{\text{migration from outside } \mathcal{D}}. \quad (26)$$

The language state of  $\mathcal{D}$  then evolves (assuming  $\mathbf{h} = \mathbf{1}$ ) as follows

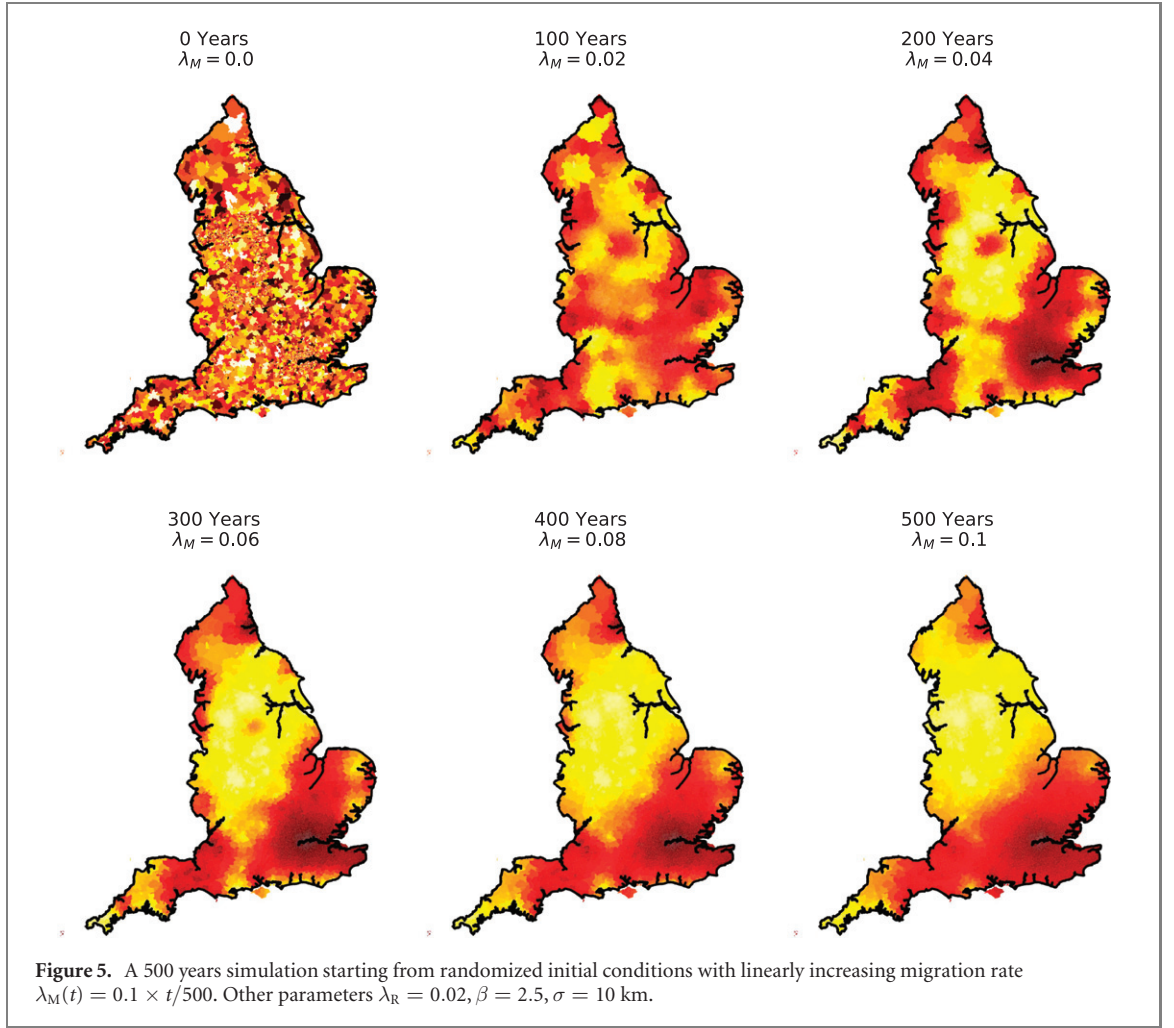
$$\Delta f_1 = \underbrace{\lambda_R \left[ \frac{f_1^\beta}{f_1^\beta + (1 - f_1)^\beta} - f_1 \right]}_{\text{local conformity}} + \underbrace{\epsilon(r)\lambda_M [\phi_1 - f_1]}_{\text{long range levelling}}. \quad (27)$$

The first term generates conformity to the majority variant within  $\mathcal{D}$ . The second generates levelling toward the external language state. Suppose that most speakers in  $\mathcal{D}$  use variant 1, which distinguishes them from the nation as a whole, where variant 2 is used (so  $\phi_1 \approx 0$ ). The equilibria ( $\Delta f_1 = 0$ ) of (27) determine the conditions under which  $\mathcal{D}$  will retain its distinctive feature. Informally, we look for the values of  $f_1$  for which the rates of conformity and levelling are in balance

$$|\text{local conformity}| = |\text{national levelling}|. \quad (28)$$

When  $\phi_1 = 0$ , this condition may be written

$$\frac{f_1^\beta}{f_1^\beta + (1 - f_1)^\beta} = \left( 1 + \frac{\epsilon(r)\lambda_M}{\lambda_R} \right) f_1. \quad (29)$$



An upper bound on the migration rate which allows variant 1 to persist within  $\mathcal{D}$  is obtained by considering solutions in the infinite conformity limit  $\beta \rightarrow \infty$  where learners always select the majority local variant. In this case the left-hand side of (29) is a unit step at  $f_1 = \frac{1}{2}$ , and there is a stable equilibrium  $f_1^* > \frac{1}{2}$  provided

$$\frac{\epsilon(r)\lambda_M}{\lambda_R} < 1. \quad (30)$$

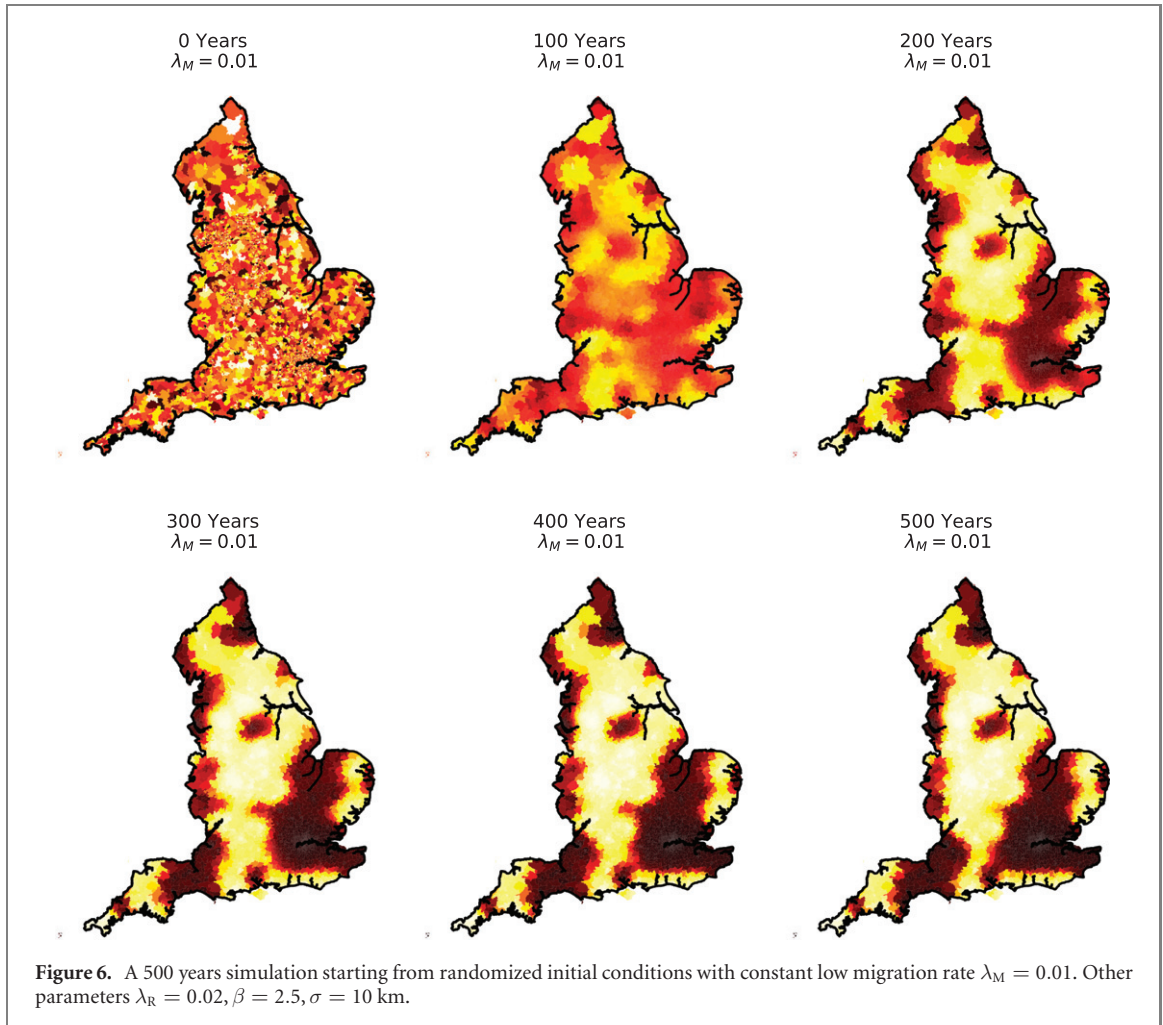
This is an intuitively reasonable condition; it says that a local distinctive feature can in principle survive provided the rate of inward migration does not exceed the rate at which young indigenous speakers replace dying adults. We estimate that during the 20th century,  $\lambda_M \approx 0.1$ . There is some flexibility in the way we interpret  $\lambda_R$ . If we consider speakers to be members of the speech community from birth, then  $\lambda_R^{-1}$  is simply life expectancy. However, if speakers are only influential once older, and cease to be so when very old, then  $\lambda_R^{-1}$  is the average of this time interval. As noted in section 3.3, we have set  $\lambda_R = 0.02$ . What determines domain stability is not however the absolute values of the migration and replacement rates. It is their ratio

$$\rho \triangleq \frac{\lambda_M}{\lambda_R}. \quad (31)$$

For given  $\rho$ , the stability of  $\mathcal{D}$  can be computed for any value of  $\beta$  and  $r$  by checking that the dynamics (27) has a stable fixed point  $f_1^* > \phi_1$ . The phase diagram in figure 4 shows that for lower migration–replacement ratios smaller domains can persist for lower values of conformity. In systems where the migration–replacement ratio grows slowly, such as in the English dialect area in previous centuries, we would expect smaller dialect areas to disappear as the phase boundary shifted.

#### 4.2. Competition between migration and conformity in the English dialect area

We consider a hypothetical period of 500 years, beginning with a domain in a completely randomised state. This state is constructed by setting the initial frequency  $f_1(\mathbf{r}, 0)$  of each cell to be an independent uniform random variable on  $[0, 1]$ . We allow the migration rate to be a linearly increasing function of time, starting



with zero migration and reaching late 20th century levels by end of the period

$$\lambda_M(t) = 0.1 \times \frac{t}{500}. \quad (32)$$

Within a few decades, recognisable dialect domains emerge (figure 5). These are analogous to domains in ferro-magnetic materials which form because neighbouring atoms tend to align their magnetic directions. Domain boundaries evolve as if they feel a form of surface tension [101] which penalises curvature. In physical systems, domains gradually expand as their boundaries straighten. The analogy between these magnetic domain boundaries, and isoglosses, was explored in detail in [7, 29], including the effects of non-uniform population distribution. It was shown in [7] that concentrated human settlements (villages, towns, cities) balance the surface tension effect by inducing isogloss curvature, implying that stable isogloss shapes are partially predictable from population distributions. At low migration levels, competition between isogloss smoothing, induced by surface tension, and curvature, induced by population distribution (as well as other natural barriers) leads to stable geographical distributions. In figure 5 the map obtained after 200 years would be stable in the absence of increasing migration. This stability is revealed in figure 6, which shows how the system evolves starting from an initial condition which is an exact copy of that in figure 5, but under conditions of consistently low migration ( $\rho = \frac{1}{2}$ ). This shows that the pattern reached after 200 years remains stable for the next 300 years, and includes a number of isolated dialect areas with small escape radii. Returning to figure 5 we see that as migration increases these small dialect areas disappear, being replaced with the variant which is in the majority further afield. Eventually the system reaches a state observed in both the SED and EDA surveys, which consists of a major isogloss bisecting England. The stable north east dialect centred on Tyneside is also seen in survey data. The major Wash–Severn isogloss is particularly stable because neither variant is in the majority in the country as a whole, and because it bisects the country via one of the shortest routes. A detailed mathematical explanation of these effects, which are related to continuum percolation, is given in [7, 29]. The north east dialect is stable in this simulation because of its status as a dense but geographically isolated population centre.



## 5. Inference for English dialects

### 5.1. Methodology

We now explore how our model predicts the evolution of linguistic variables which have been recorded in both the SED and the EDA. These are listed in table 1. For each variable we evolve our model  $T = 100$  years forward in time using its SED state as our initial condition. To generate this initial condition from raw survey data, we set variant frequencies in each model cell (MSOA) to match the variant frequencies from the nearest SED survey location. If multiple variants were recorded in this location, they are assumed to all have equal frequency. This process yields a tessellation of England into single and shared-variant regions. Relative frequencies in each cell are then replaced with average frequencies over its 20 nearest neighbours (mean separation 5.4 km). This amounts to a form of smooth interpolation between relative frequencies observed at SED survey points, which have typical separation  $\approx 20$  km (calculated as square root of mean land area per location). The resulting distribution has smooth transitions with widths typically 10–30 km, consistent with English transition zones discussed in reference [18]. While we would not expect our results to be sensitive to the number of nearest neighbours used in the smoothing (interpolation) step, we have avoided smoothing to a level where frequency values at SED locations would be changed substantially in transition regions (e.g. if the smoothing range approached the typical separation of survey locations).

We consider both bias free ( $\mathbf{h} = \mathbf{1}$ ) and biased evolution. We allow for the possibility that biases may be different in different parts of the country by introducing a spatially varying field

$$\mathbf{h}(\mathbf{r}) = (h_1(\mathbf{r}), \dots, h_q(\mathbf{r}))^T, \quad (33)$$

where  $h_k(\mathbf{r})$  is the bias on variant  $k$  in cell  $\mathbf{r}$ . We do not allow the bias field to depend on time because we have only initial and final conditions, and cannot therefore calibrate time dependence. We wish to fit  $\mathbf{h}(\mathbf{r})$  so that the final state of our simulation, initialised with an SED variable, closely matches the corresponding EDA data. This is achieved by iterating the following *learning step*, which increases or decreases local biases on variants which are respectively under- or over-represented in the model, when compared to the EDA

$$\hat{\mathbf{h}}_{n+1}(\mathbf{r}) = \eta(\mathbf{f}_{\text{EDA}}(\mathbf{r}) - \mathbf{f}_n(\mathbf{r}, T)) + \theta(\mathbf{1} - \mathbf{h}_n(\mathbf{r})). \quad (34)$$

Here  $\mathbf{h}_n(\mathbf{r})$  is the  $n$ th iteration of our bias estimate and  $\mathbf{f}_n(\mathbf{r}, T)$  is the frequency distribution obtained using this estimate. The *learning rate*,  $\eta$ , controls how rapidly adjustments are made, and the *reversion rate*,  $\theta$ , is a regularization parameter that prevents runaway bias increases, and maintains  $\mathbf{h} = \mathbf{1}$  as the ‘no-bias’ reference point so that calibrations for different variables can be compared. We use the initial condition  $\mathbf{h}_0(\mathbf{r}) = \mathbf{1}$ .

To avoid over-fitting we introduce a smoothing factor,  $\sigma_s$ , which interpolates between independent variation in bias between cells ( $\sigma_s = 0$ ) and constant bias in all cells ( $\sigma_s \rightarrow \infty$ ). The quantity  $\hat{\mathbf{h}}_{n+1}(\mathbf{r})$  is the new bias estimate *before* spatial smoothing (indicated by the hat symbol). Having applied the learning step (34), we apply a *smoothing step* using the interaction matrix (2) with interaction range  $\sigma_s$ ,

$$\mathbf{h}_{n+1}(\mathbf{r}) = \sum_{\mathbf{r}'} W_{\sigma_s}(\mathbf{r}, \mathbf{r}') \hat{\mathbf{h}}_{n+1}(\mathbf{r}'). \quad (35)$$

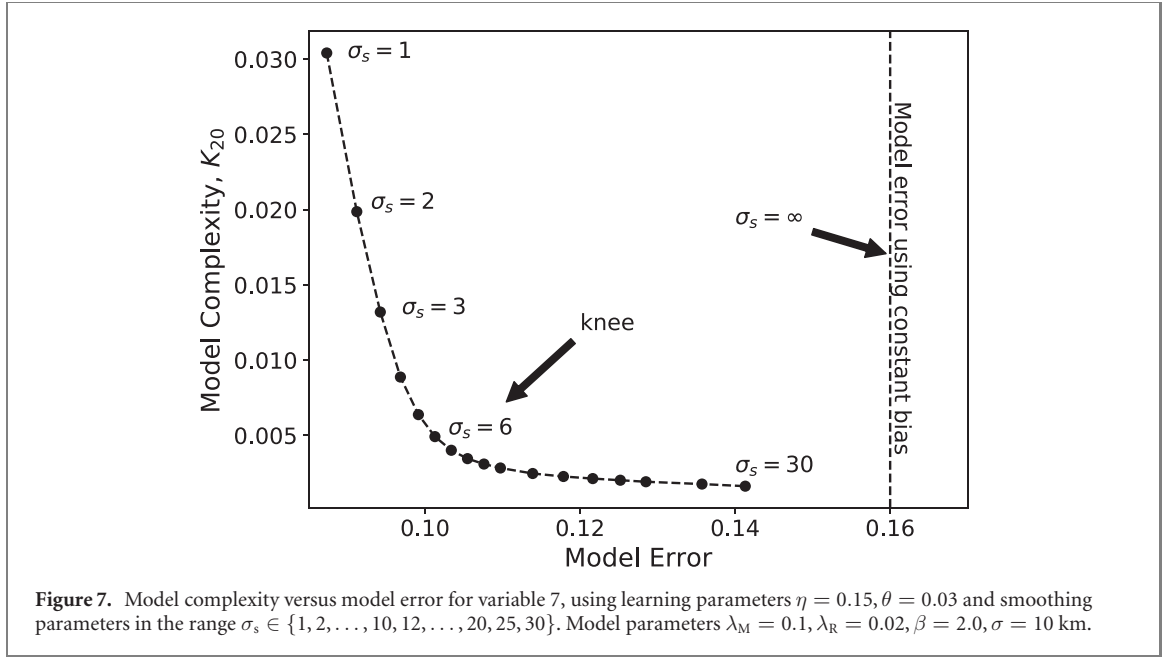
This two-step iterative estimation method is continued until the bias field stabilizes. The value of  $\sigma_s$  is determined by optimising the trade off between model complexity and model error. We measure complexity,  $K_N$ , as the average  $\ell_1$  norm of the difference between the bias in each cell, and the average,  $\bar{\mathbf{h}}_N(\mathbf{r})$ , of its  $N$  nearest neighbours

$$K_N = \sum_{\mathbf{r}} \|\mathbf{h}(\mathbf{r}) - \bar{\mathbf{h}}_N(\mathbf{r})\|_1. \quad (36)$$

This measure approximates the magnitude of the second spatial derivative of the bias field, averaged over the whole system. The number of neighbour cells ( $N$ ) is chosen to be large enough to smooth out inhomogeneity in the pattern of cell centroids, but small enough so that shorter range fluctuations in the bias field are not ignored. We take  $N = 20$ , corresponding to an average neighbour distance of 5.4 km, on the assumption that repeated application of the smoothing step will ensure that fluctuations on shorter scales are small. Complexity will be high for fields with large variations in bias over short distances throughout the system, and low for fields which are approximately constant over large geographical areas. Higher learning rates will generate bigger local changes at each step, but these are counteracted by increased smoothing range. Reducing the learning rate to  $\eta' = \eta/k$ ,  $k \in \mathbb{N}$ , is approximately equivalent to applying  $k$  smoothing steps between each learning step, or increasing the smoothing parameter to  $\sigma'_s = \sqrt{k}\sigma_s$ , so we can explore the behaviour of our calibration method by fixing  $\eta$  and varying  $\sigma_s$ . Figure 7 shows the relationship between model complexity and model error using variable 7 as an example. For high smoothing parameter ( $\sigma_s = 30$  km) the error approaches that of the maximally parsimonious model, where bias is constant in space. Reducing  $\sigma_s$  improves the model error, until

Table 1. Questions, variables and variants for the EDA (and SED). All phonetic variants are written using the International Phonetic Alphabet [102]. ME = Middle English.

Question in the EDA	Variable	Variants
1. The word ‘tongue’ can end in two different ways. Which do you use?	Coalescence of [ŋg]	1. [ŋg], 2. [ŋ]
2. Which pronunciation of ‘new’ is the most similar to your own?	Yod-dropping	1. no [j], 2. [j]
3. How do you pronounce the word ‘last’?	TRAP–BATH split	1. [ɑ:], 2. [a:], [æ:] 3. [a], [æ]
4. In the word ‘butter’, I pronounce the letter ‘u’ as...	FOOT–STRUT split	1. [ʊ], 2. [ʌ]
5. Do you pronounce the ‘r’ in ‘arm’?	Rhoticity	1. no /r/, 2. /r/
6. How do you pronounce the word ‘three’?	Realisation of word-initial/θ/	1. [θ], 2. [d], [d], 3. [f], 4. [t]
7. Do you pronounce the ‘r’ before the ‘-ing’ in ‘thawing’?	Intrusive /r/	1. /r/ 2. no /r/
8. What is the season that follows summer?	Autumn lexical item	1. autumn, 2. backend, 3. fall
9. How do you pronounce the ‘l’ in ‘shelf’?	l-vocalisation	1. [ɫ], 2. [u], [ʊ], 3. [l]
10. What do you call a piece of wood stuck under your skin?	Splinter lexical item	1. shiver, 2. sliver, 3. speel, 4. spelk, 5. spell, 6. spile, 7. spill, 8. splint, 9. splinter, 10. spool
11. How do you pronounce the word ‘room’?	Realisation of ME /o:/ in ‘room’	1. [Y:], [Y], 2. [u:], 3. [ʊ]
12. How do you pronounce the word ‘chicken’?	Weak vowel merger	1. [ə], [ʌ], 2. [ɪ]
13. How pronounce the word ‘night’?	Realisation of ME /ix/	1. [ɑɪ], 2. [ɑ:], 3. [ɔɪ], [ɒɪ], 4. [ɛɪ], 5. [ʌɪ], [əɪ], 6. [aɪ], [æɪ], 7. [a:], [æ:], 8. [i:]
14. He wasn’t careful with the knife and managed to cut...	Initial element in 3sg m. reflexive pronoun	1. himself, 2. hisself
15. If it belongs to a woman its...	3sg f. possessive pronoun	1. hern, 2. hers
16. Do you pronounce the ‘h’ in ‘hands’?	h-dropping	1. /h/, 2. no /h/
17. If you’d like someone to pass you something, would you say:	Dative alternation	1. give it me, 2. give it to me, 3. give me it
18. How do you pronounce the word ‘off’?	LOT-CLOTH split	1. [ɔ:], 2. [ɑ:], 3. [ɒ]
19. How do you pronounce the word ‘house’?	Realisation of ME /u:/	1. [əu], [ʌu], [œʏ], 2. [ɛ:], 3. [ɛu], 4. [aɪ], 5. [a:], 6. [æu], [æʏ], 7. [au], 8. [u:]
20. How do you pronounce the word ‘bacon’?	Realisation of ME /a:/	1. [ɪə], [ɪa], 2. [æɪ], 3. [e:], [eə], 4. [ei], [ɛi], [ɛ:], [ɛə]
21. How do you pronounce the word ‘five’?	Realisation of ME /i:/	1. [ɑɪ], 2. [ɑ:], 3. [ɔɪ], 4. [ɛɪ], 5. [ʌɪ], [əɪ], 6. [aɪ], [æɪ], 7. [a:], [æ:]
22. How do you pronounce the last sound of ‘bit’ when saying ‘a bit of’?	Coda /t/	1. [ʔ], 2. [d], [r], 3. [t]
23. Fill in the gap: every day on her walk past the lake, she ... the ducks.	Habitual present 3sg.	1. do feed, 2. feed, 3. feeds
24. How do you pronounce the word ‘happy’?	happyY tensing	1. [ə], 2. [ɪ], 3. [e], 4. [i]
25. An animal that carries its house on its back is a...	Snail lexical item	1. dod-man, 2. hodmedod, hoddly-dod, hoddly-doddy, 3. snail



we reach an optimal region ( $\sigma_s \approx 6$  km), the ‘knee’ point, with low model errors and minimal complexity. Further increasing complexity has minimal effect on error. We take this knee as our heuristic for determining the optimal level of complexity because this is the simplest explanation that fits the data (Occam’s razor [103]). Exploration of the error vs complexity trade-off for other questions reveals similarly located knee points. For consistency we use  $\sigma_s = 6$  km for all questions (results with  $\sigma_s = 10$  km are shown in supplementary material (<https://stacks.iop.org/JPCOMPLEX/2/035018/mmedia>) [104]). For comparison, we also calibrate with very high smoothing  $\sigma_s = 150$  km in the case  $\beta = 2$ , which produces an approximately constant bias field. We select learning parameters  $\eta = 0.15, \theta = 0.03$  to allow convergence within a feasible time frame, while maintaining stability in the iteration.

We measure the difference between modelled and observed variant distributions at time  $T$  using the *total variation distance* [105] between the variant frequency distributions in each cell, averaged over all cells

$$\text{error} \triangleq \frac{1}{2L} \sum_{\mathbf{r}} \|\mathbf{f}_{\text{model}}(\mathbf{r}) - \mathbf{f}_{\text{EDA}}(\mathbf{r})\|_1 \in [0, 1], \quad (37)$$

where  $\|\cdot\|_1$  is the  $\ell_1$  norm. The factor  $(2L)^{-1}$  ensures that the maximum error is 1. In this extreme case no speaker in any cell uses a variant observed in that cell in the survey.

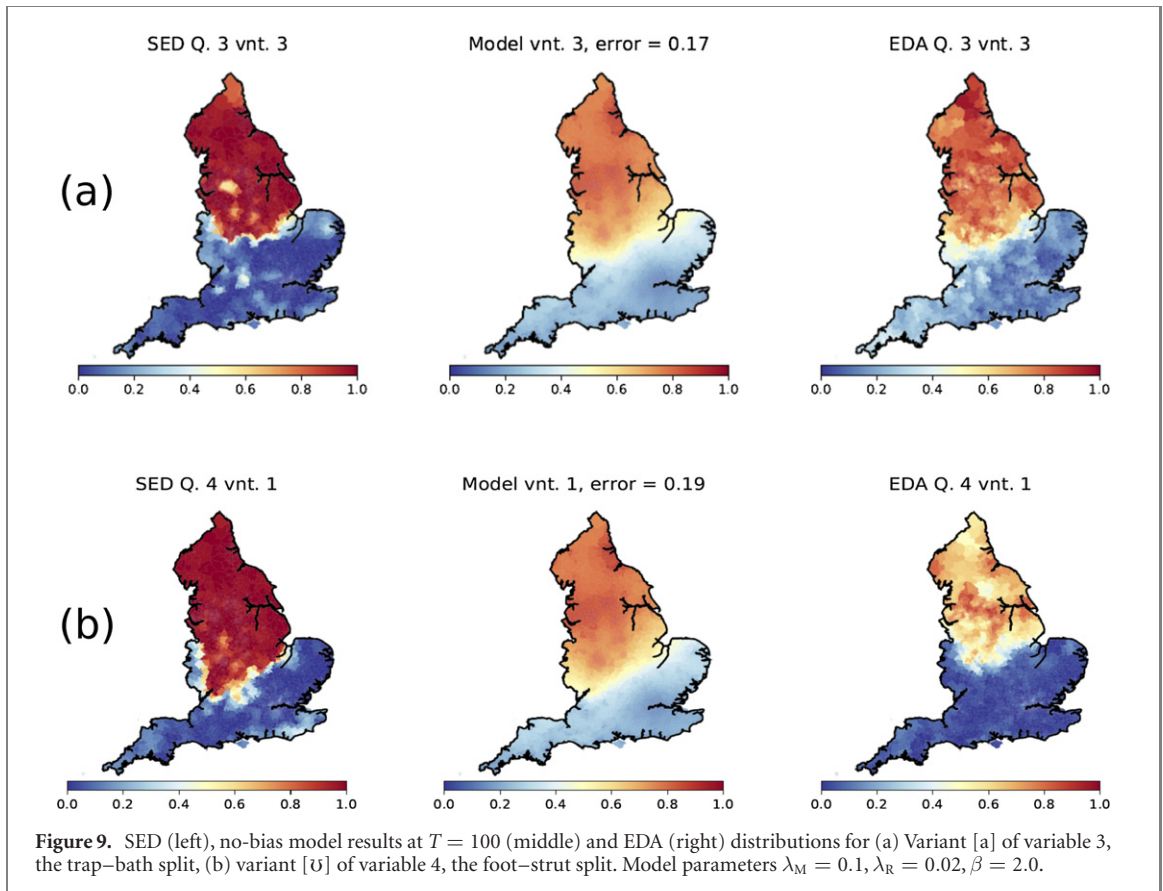
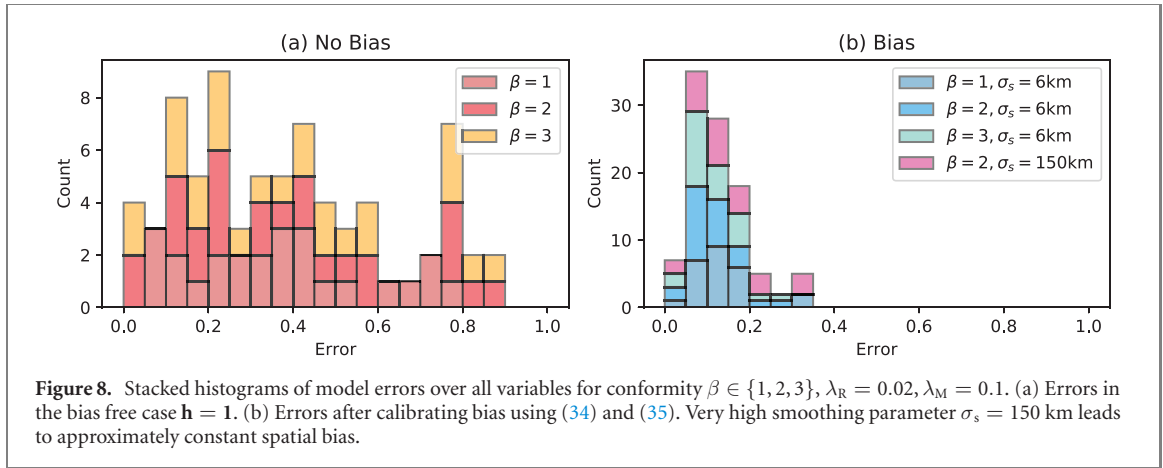
The model is implemented in vectorized Python using sparse arrays which accelerate the calculation of the community frequency vector. The number of bias field iterations is set between 25 and 35, sufficient for the error (37) to stabilize. To calibrate a single question using a single set of model parameters takes approximately 5–10 min on a modest laptop (1.7 GHz, 4 cores). For much larger systems than ours, there is potential for the migration matrix to create a bottleneck, requiring further optimization.

## 5.2. Bias free results

We begin by considering the bias free case ( $\mathbf{h} = \mathbf{1}$ ) where the only factors driving evolution are migration, daily movement and conformity. Figure 8(a) shows the range of model errors in this case. In approximately 25% of cases the model generates a final distribution with error ( $< 0.2$ ) comparable with errors obtained by fitting a bias field (figure 8(b)). These are variants for which spatial process and conformity play a substantial role in explaining their evolution. Many exhibit one of two phenomena seen in two dimensional coarsening systems studied by physicists [101]: stripe states, and shrinking droplets.

### 5.2.1. Stripe states

In the bias free case, provided  $\beta > 1$ , our model behaves like a subcritical coarse grained two-dimensional ferromagnet [101, 106]. The condition  $\beta > 1$  allows interfaces to form between single variant domains, analogous to domains of uniform spin alignment in magnetic systems. These interfaces feel a form of surface tension, and in finite systems, stable configurations often emerge where interfaces connect opposite system boundaries [106], creating *stripes*. Variant [a] of variable 3 in table 1 illustrates this phenomenon (figure 9(a)). Here, we see that the geographical boundary between those who pronounce the vowel in the bath lexical set

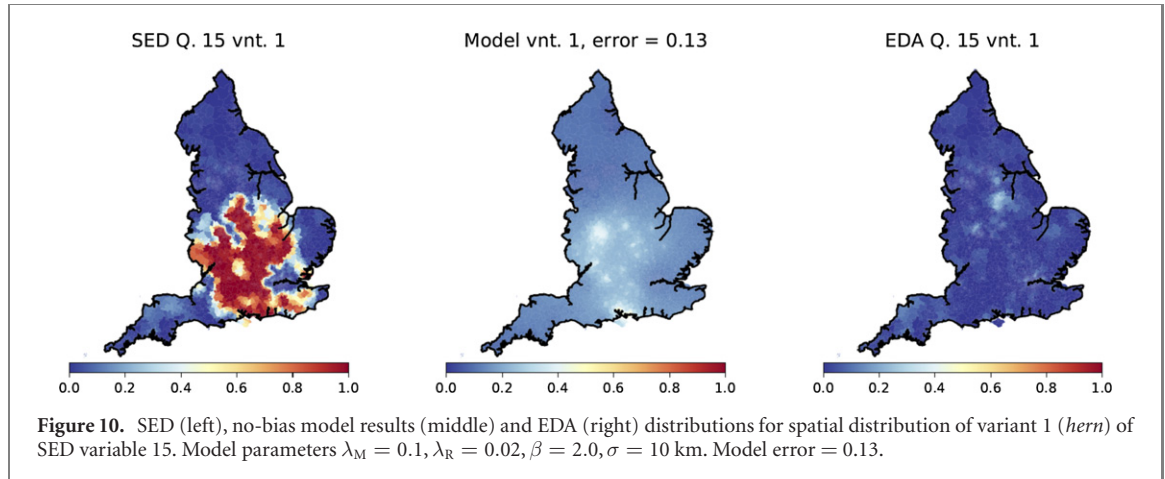


[107] (*last, staff, brass, ...*) in the same way as vowel as the trap set (*tap, back, badge, ...*), from those who do not, has been stable over the 20th century. Note that *lexical sets* (introduced by Wells [107]) are groups of words defined by some common phonological feature. The variant [a] (the vowel in northern ‘bath’) dominates in the northern region, where the vowels in TRAP and BATH do not differ. This same stability is predicted by the model, and derives from the surface tension effect which causes interfaces to find the shortest routes across the system, which often begin in boundary indentations [7] (the Wash in figures 9(a) and (b)). One important difference between our model and simple magnetic systems is migration, which induces effective long-range interactions equivalent to local biases in favour of the national majority variant. Because the trap–bath split divides the country into approximately equal parts, these biases are close to zero, stabilizing the boundary.

Another example of this effect, shown in figure 9(b), is the distribution of the variant [ʊ] of variable 4. This is the northern English form of the stressed vowel in ‘butter’ which belongs to the Wells’ lexical set strut (*cup, suck, ...*). Speakers from the north typically use the same vowel sound for the lexical set foot (*put, bush*). Although the no-bias model captures the broad overall pattern of foot–strut, we can see from the EDA distribution that the northern form is in fact receding, which indicates a bias in favour of the southern pronunciation,

**Table 2.** Mean errors for  $\beta \in \{1, 2, 3\}$ . Here  $\varepsilon_{\text{bias}}$  and  $\varepsilon$  denote errors in the bias and no-bias models respectively. The shorthand  $\langle \varepsilon | \varepsilon < C \rangle$  denotes mean error conditional on  $\varepsilon < C$ . Counts give the numbers of questions satisfying the condition. Note that  $C = 0.17$  is one standard deviation above the mean error in the  $\beta = 2$  bias model, and  $C = 0.22$  is two standard deviations above. For comparison, the mean error with  $\beta = 2$  and spatially constant bias is  $0.15 \pm 0.09$ .

$\beta$	$\langle \varepsilon_{\text{bias}} \rangle \pm \text{std.dev.}$	$\langle \varepsilon   \varepsilon < 0.17 \rangle$	Count( $\varepsilon < 0.17$ )	$\langle \varepsilon   \varepsilon < 0.22 \rangle$	Count( $\varepsilon < 0.22$ )
1	$0.133 \pm 0.070$	0.105	6	0.121	7
2	$0.109 \pm 0.058$	0.087	5	0.123	8
3	$0.112 \pm 0.061$	0.089	5	0.126	8



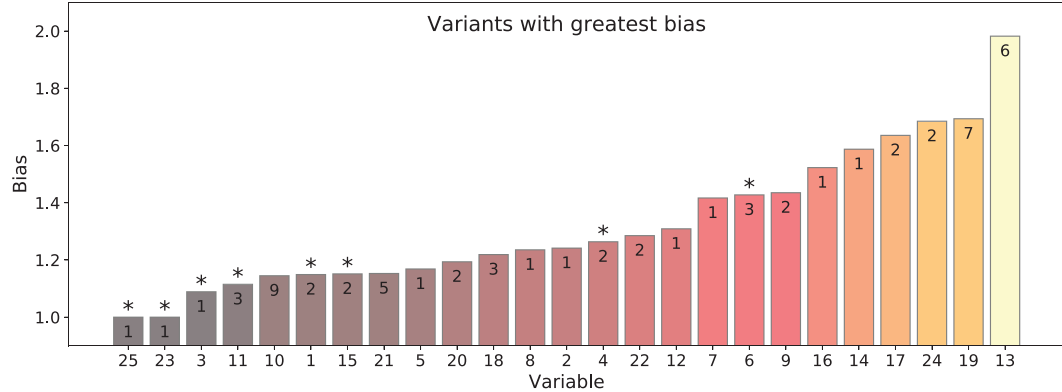
[ $\Delta$ ]. Further evidence of such a bias, is that the model predicts levelling of both variants, when in fact the southern form remains dominant in the EDA.

The effect of increasing the conformity number is to make single variant regions increasingly pure, boundaries sharper, and more strongly affected by variations in population distribution. In this paper, for every variable, we have run the model with  $\beta \in \{1, 2, 3\}$ . Mean errors in the biased and no-bias model are summarized in table 2. For the no-bias model we compute average errors excluding cases where these exceed a given threshold,  $C$ , allowing us to consider only cases where the model provides a plausible description of changes. In the biased model we find that conformity generates a better fit than no conformity (a paired  $t$ -test with null hypothesis of equal mean error between  $\beta = 1$  and  $\beta = 2$  yields  $p$ -value 0.0016). There are small differences ( $\lesssim 0.01$ ) between the cases  $\beta = 2$  and  $\beta = 3$ . In the no-bias model, when errors are within one standard deviation of the biased model error, the conformity model performs marginally better than the conformity free case. If we raise the error threshold to two standard deviations the results are equivocal. Table 2 therefore provides only limited evidence that  $\beta > 1$  in the no-bias case. We suggest that  $\beta > 1$  is more plausible because in the long term, linguistic boundaries could not spontaneously emerge or be maintained if  $\beta = 1$ . Moreover, without conformity the unbiased dynamics simply mixes variants, dissolving boundaries, but leaving total variant frequencies over the system as a whole unchanged, meaning that levelling is impossible. Previous modelling work [31] also indicates that the SED distributions are more likely to have been generated by a model with conformity. Finally, we note that the substantial linguistic changes observed during our period of interest may be driven by a wide range of factors, of which conformity is just one, making its effects harder to discern. A complete set of maps for all variants of all variables for the SED, EDA, no-bias, and bias models (bias fields included) for  $\beta = 2$ , is given in the supplemental material for this paper. For the remainder of the paper we focus on the case  $\beta = 2$ , and consider spatial varying bias ( $\sigma_s = 6$  km), approximately constant bias ( $\sigma_s = 150$  km) and no bias.

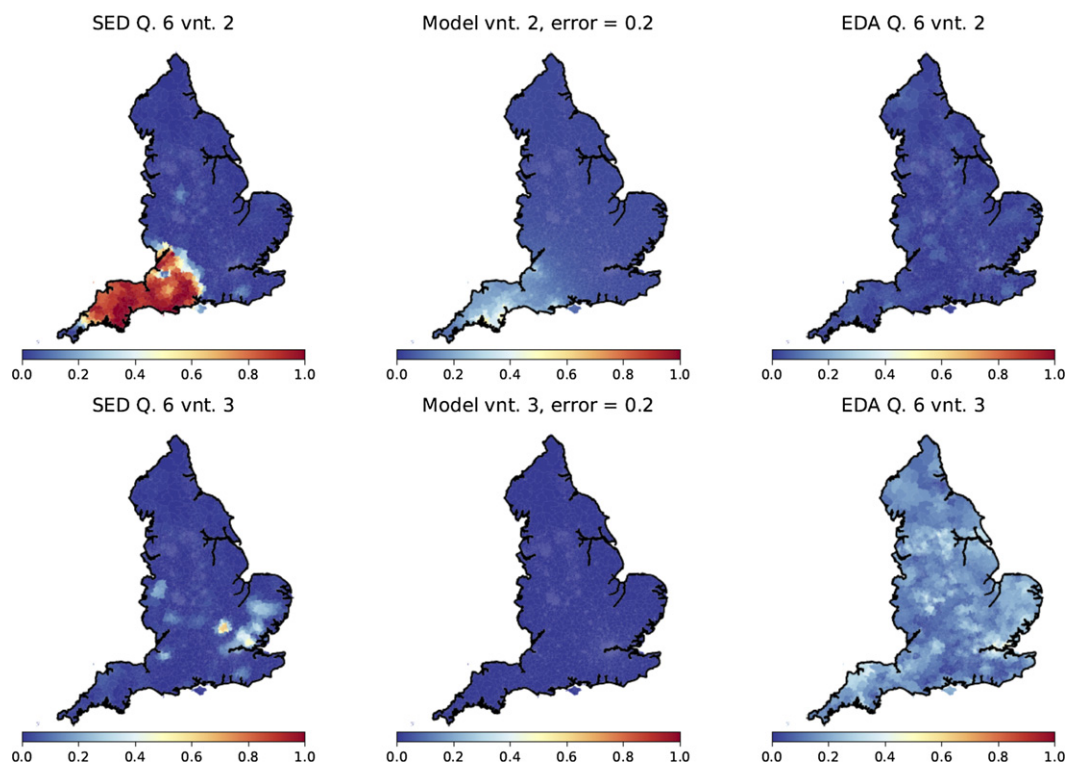
### 5.2.2. Shrinking droplets

Another phenomenon seen in ferromagnetic interfaces, and in our model, is the *shrinking droplet*. In two dimensions a droplet is simply a closed curve, evolving under surface tension, which acts to shorten its length. As a result, droplets become more circular and eventually shrink to nothing. In our model, droplets of variants which are in the minority are further reduced by migration. As an example, consider the form of the 3sg feminine possessive pronoun, *hers* or *hern* (variable 15, figure 10). *Hern* is a minority variant in the SED (29% vs 71% nationally), surrounded by regions in which *hers* is dominant. Without conformity, the mixing effect of migration would move the system as a whole to a spatially uniform state with all cells having frequency vector





**Figure 11.** Variants with the maximum average bias per cell (given by bar height) in the bias model with parameters  $\lambda_M = 0.1$ ,  $\lambda_R = 0.02$ ,  $\beta = 2.0$ . Each bar is labelled with the variant that experiences the maximum bias. Starred bars indicate variables for which the no-bias model with  $\beta = 2$  generates an error  $< 0.2$ .

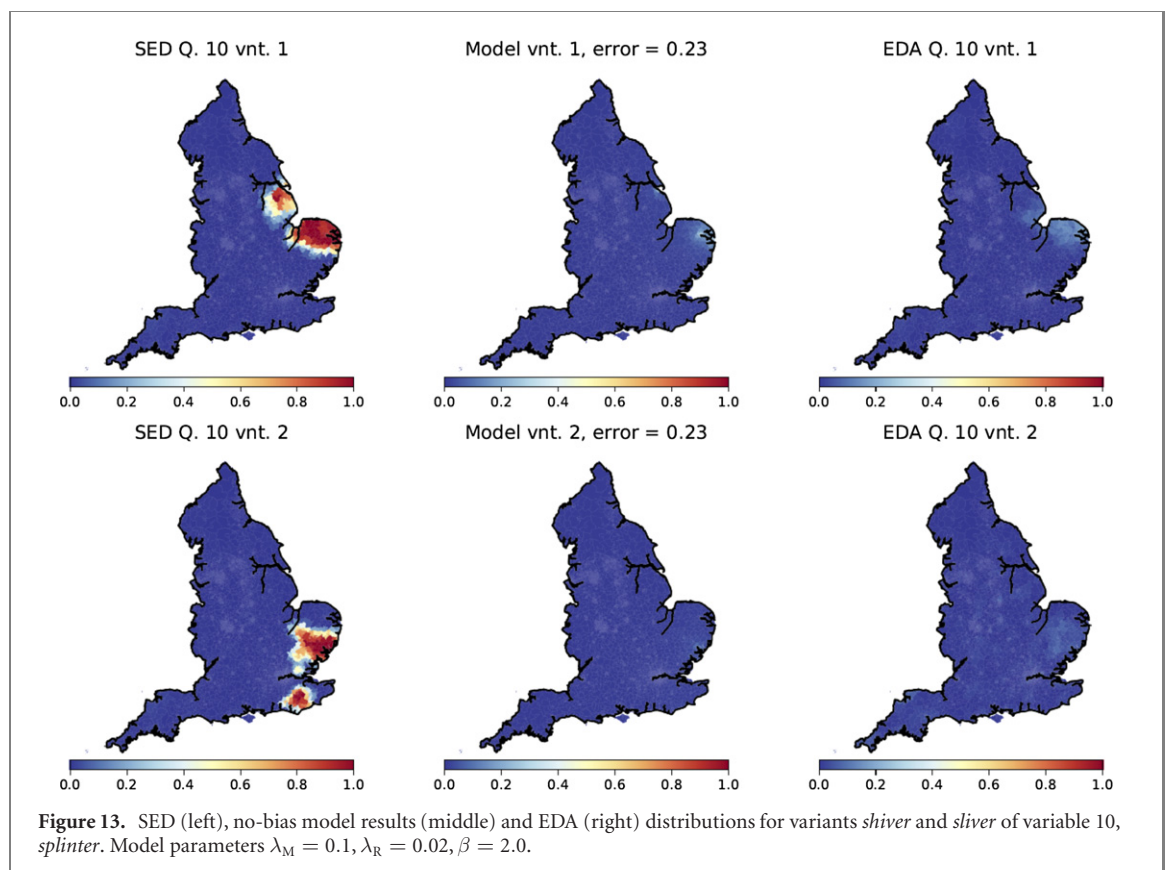


**Figure 12.** SED (left), no-bias model results (middle) and EDA (right) distributions for variant 2 ([d.i:ɪ]) and variant 3 ([f.i:ɪ]) of variable 6: pronunciation of 'three'. Model parameters  $\lambda_M = 0.1$ ,  $\lambda_R = 0.02$ ,  $\beta = 2.0$ .

$\mathbf{f} = (0.29, 0.71)$ . With conformity, surface tension acting at the boundary of the *hern* region, plus the effects of migration, almost extinguish the *hern* variant within 100 years. Examples of droplet shrinkage from a different class of linguistic feature are found in the reflexes of middle English /u:/ (variable 19), where minority variants such as [ɛ:] and [aɪ] are predicted to undergo near-complete levelling (disappearance) without bias.

### 5.3. Bias driven evolution

The evolution of many variables in table 1 may be understood as the result of combined spatial and bias effects. The primary effect of bias is to lend certain variants a special status which may have either linguistic (*internal*) or social (*external*) origin. In some cases, a bias appears to accelerate primarily spatial processes. Figure 11 shows the average bias per cell applied to the most biased variant of each variable, using  $\beta = 2$  (similar results are obtained with  $\beta = 3$ ). Variables for which the no-bias model produces an error  $< 0.2$  are starred and, with two exceptions, these variables have the smallest maximum bias per cell. Bias magnitude therefore provides an indication of the extent to which the pure spatial model fails to explain observed changes between the SED and EDA. For the more extreme of the two exceptions (variable 6, figure 12), the no-bias model correctly predicts

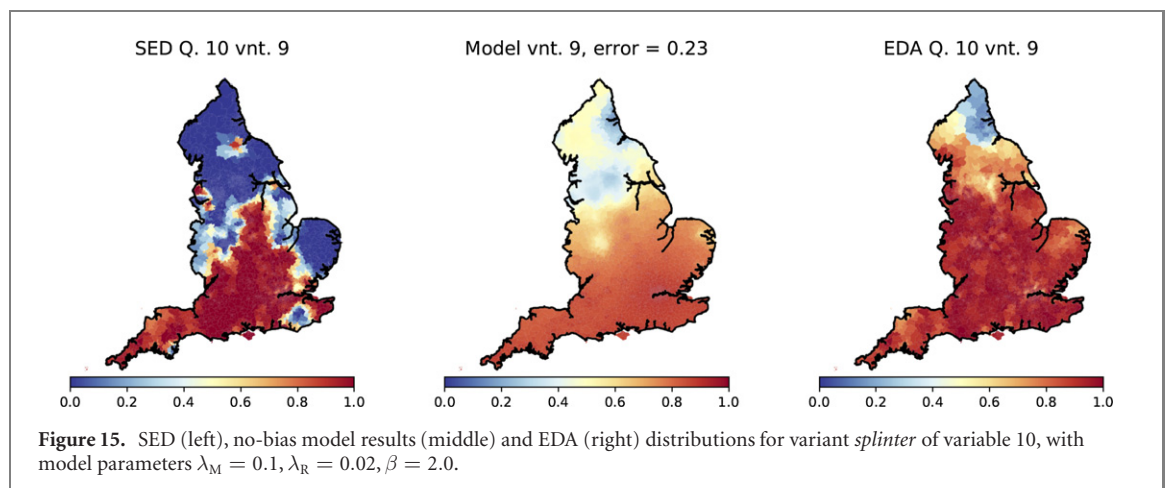
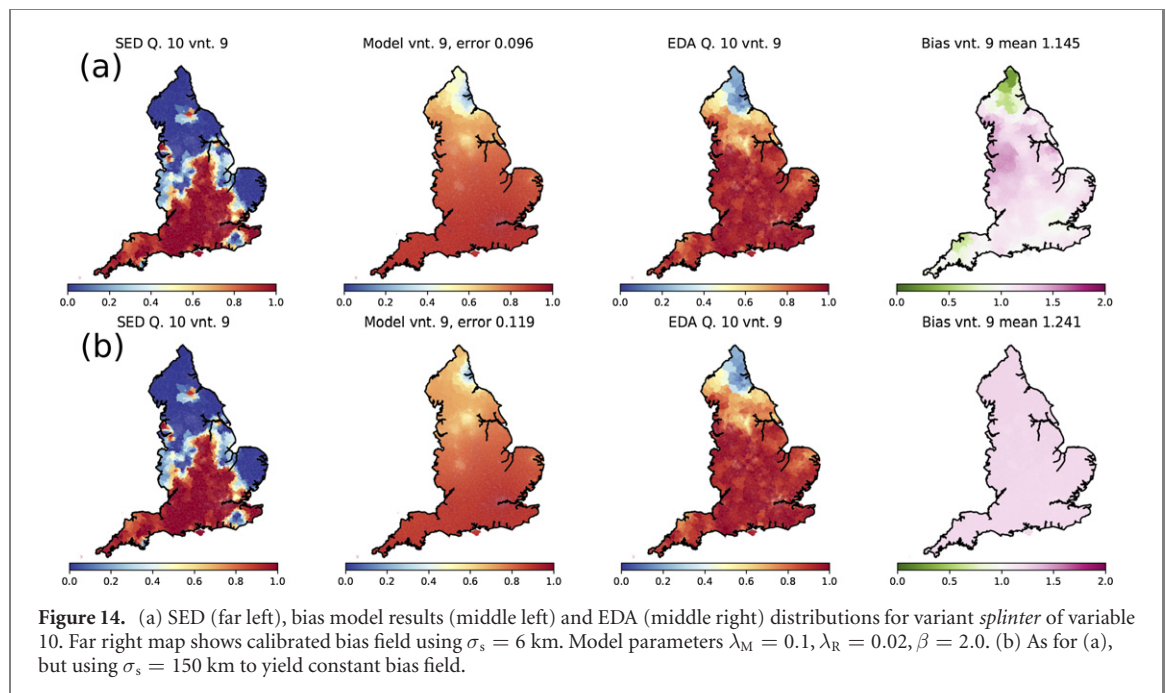


the main 20th century change: disappearance of the stopped pronunciation of ‘three’ from the west-country, but fails to predict that the highly localised Essex pronunciation ‘free’, although it remained a minority variant, would spread across the country, requiring a substantial bias in its favour, to prevent complete levelling.

### 5.3.1. Special status variants

Consider variable 10, the word for a small piece of wood stuck under the skin. Minority variants like *shiver* and *sliver* (see figure 13) are correctly predicted by the no-bias model to undergo near-total levelling. Such patterns are most typical of variables with numerous variants, many of which were historically restricted to small regions. In contrast, the present day distribution of *splinter* cannot be explained without bias (figure 14). In the SED this was the dominant southern variant, and the calibrated bias field indicates that it must have attained a special status as a standard variant in order to spread north. However, this progress was halted in far north east England, where *spelk* persists. Our spatially invariant bias field (figure 14(b)) indicates that this effect can be partially, but not completely explained by the region’s geographical isolation. According to the spatially varying bias field, *spelk* not *splinter* is the variant with special status in the north east (figure 14(a)), without which it would be dominated by *splinter*, except in the major population centre of the region (Newcastle). A similar effect is observed in variable 12, with the preservation of weak-vowel merger in the north east, when the rest of England has moved to the standard variant (no merger). The no bias model predicts the initial expansion of *splinter* due to its dominance in the densely populated south. However, its progress north is arrested by repulsion effects [7] created by the major urban conurbations of Manchester, Sheffield and Leeds (figure 15). We find that these cities play an important role in the pronunciation of ‘thawing’ (variable 7, figure 18).

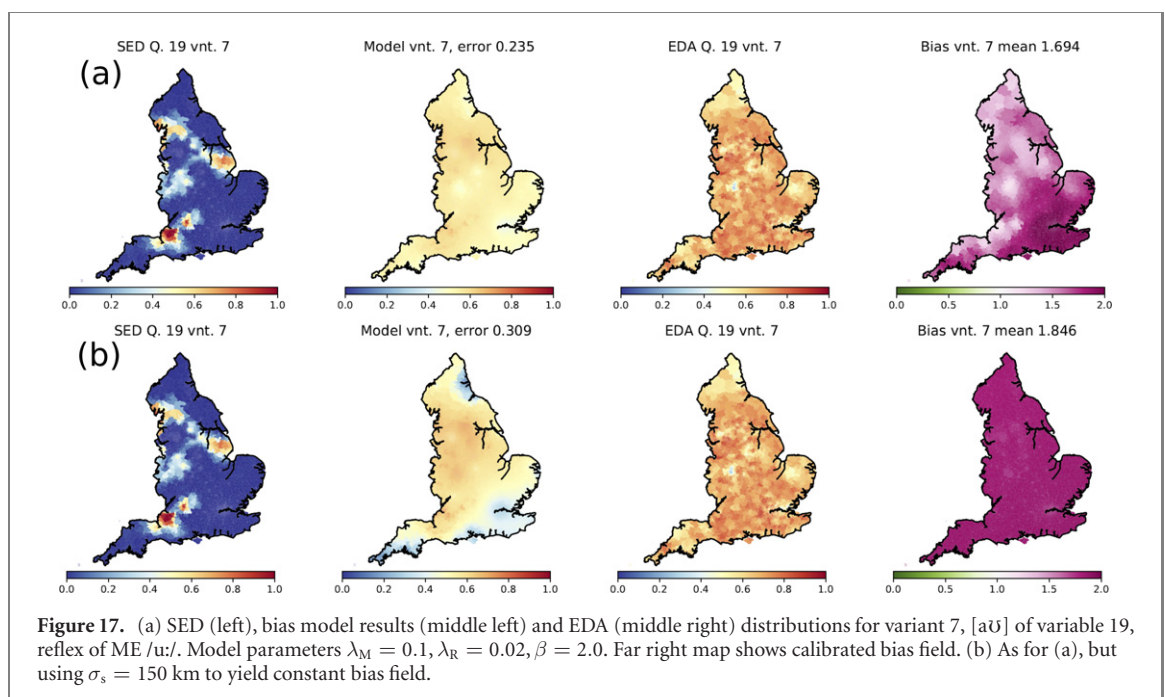
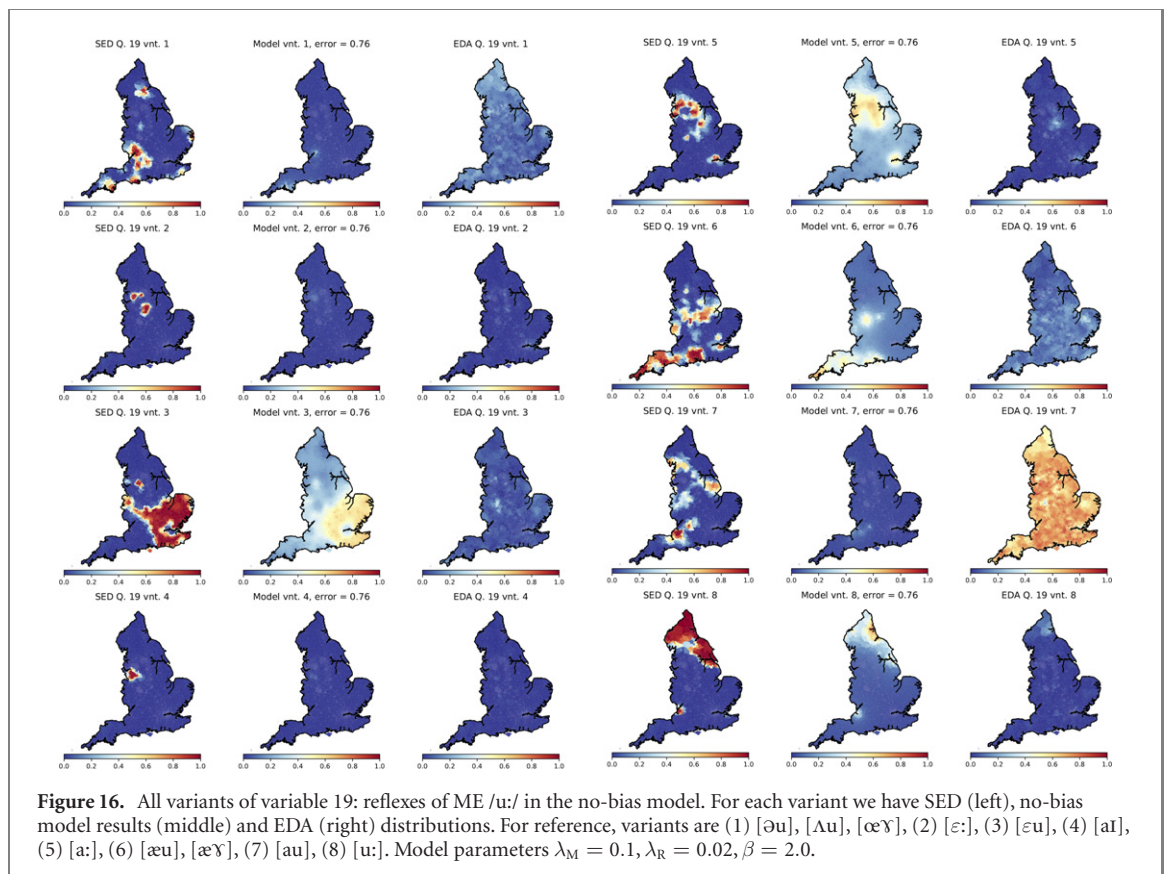
An important question is: how should the bias fields be interpreted? In the strictest sense the fields are merely model fitting parameters which measure how much asymmetry needs to be injected into the model to explain the observed changes. Where the variant symmetric model adequately explains the observed changes without the need for additional bias, we expect to see a calibrated bias field close to one. This does not mean that in reality there is no linguistic or social bias, only that conformity and migration alone are sufficient to explain the changes. For example, in figures 14 and 15 because the no bias model predicts the first part of the northerly expansion of *splinter*, minimal bias is needed in the south. The major obstacles to further expansion are the cities of Manchester and Sheffield, which consequently have high bias field. This does not imply that surrounding areas necessarily have lower bias, only that it is not needed to explain observations. The bias field is therefore most informative when viewed in conjunction with the predictions of the no-bias model (a complete catalogue is in the supplemental material [104]).



In several cases the no-bias model would fail to predict levelling at all, or would even predict levelling in the opposite direction to that observed. Here the direction of levelling is driven by bias: some variant-specific external or internal factor such as normative bias or markedness must be responsible for the rise in the successful variant. An example of the first type is variable 19 (figure 16): the reflexes of ME /u:/ (generally Wells' mouth). Although the no-bias model correctly predicts that small variants ([əʊ], [ɛ:], [aɪ]) are largely levelled out, since the major variants at the time of the SED ([u:], [ɛʊ], [æʊ], [a:] and to a lesser extent [aʊ]) were found in complex domains with none in the majority, the no-bias model incorrectly predicts a great deal of mixing without much change in relative proportions (i.e. without levelling). A powerful bias term (figure 11) is needed to predict the observed near-complete levelling to [aʊ] (see figure 17). In a different linguistic domain, we consider the initial element of the 3sg.m. reflexive pronoun (variable 14). Users of the standard variant, *him(self)*, represented a small minority in the SED and the no-bias model predicts that this variant would disappear; instead, the opposite happened, with substantial levelling towards *him(self)* in the EDA data. In the model, this requires strong bias in favour of *him(self)* (see figure 11).

### 5.3.2. Isogloss positioning

In certain cases, although the direction of change is driven by positive bias throughout the spatial domain, the model correctly predicts the new location of an isogloss as a product of coastline and population density. In figure 18, the model captures the movement northward of the intrusive r isogloss so that its current position divides the country in a line stretching from the Humber estuary to the Ribble estuary (the result of coastline indentations) with irregularly shaped differentiation within the northern area (the result of population density

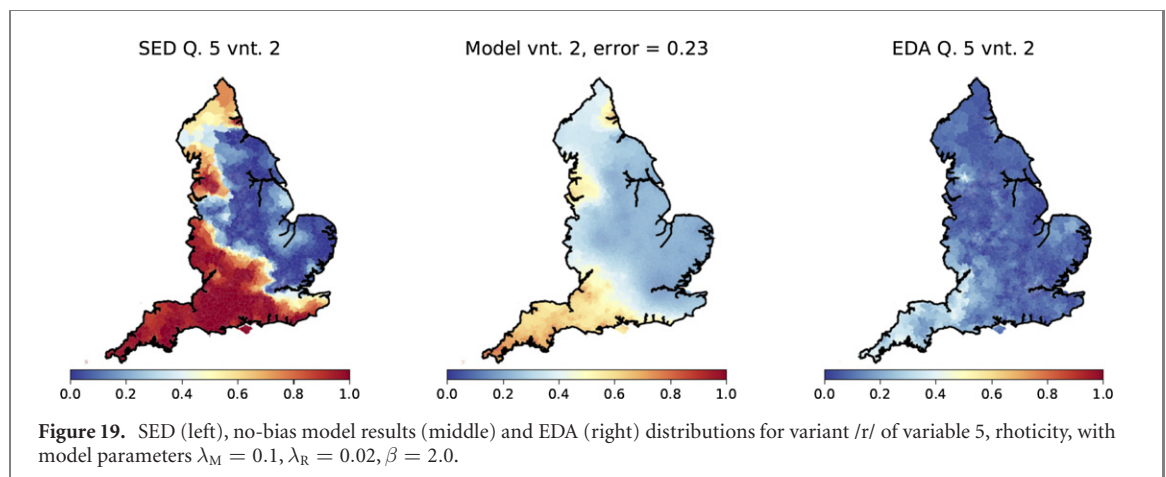
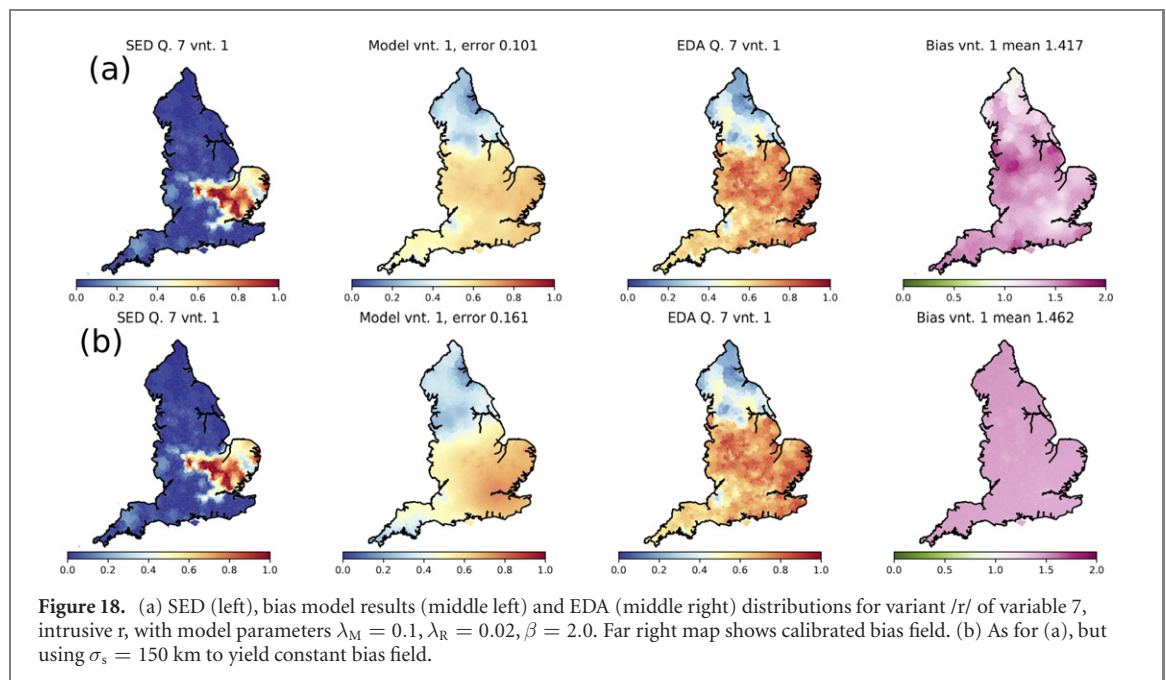


patterns). Here the spatially varying and spatially constant bias field produce very similar results (figures 18(a) and (b)) indicating that isogloss positioning is influenced in a partially predictable way by geography and population distribution, as predicted in references [7, 29]. The same isogloss appears (spuriously) in the bias free evolution of variable 10 (figure 13).

### 5.3.3. Accelerating spatially driven changes

For some variables, a pure spatial model would predict the direction of change correctly, but at a slower pace than observed: here the bias term is needed to increase the rate of change. An example is the changes affecting the circumstances in which the rhotic consonant /r/ is pronounced (variable 5): the no-bias model



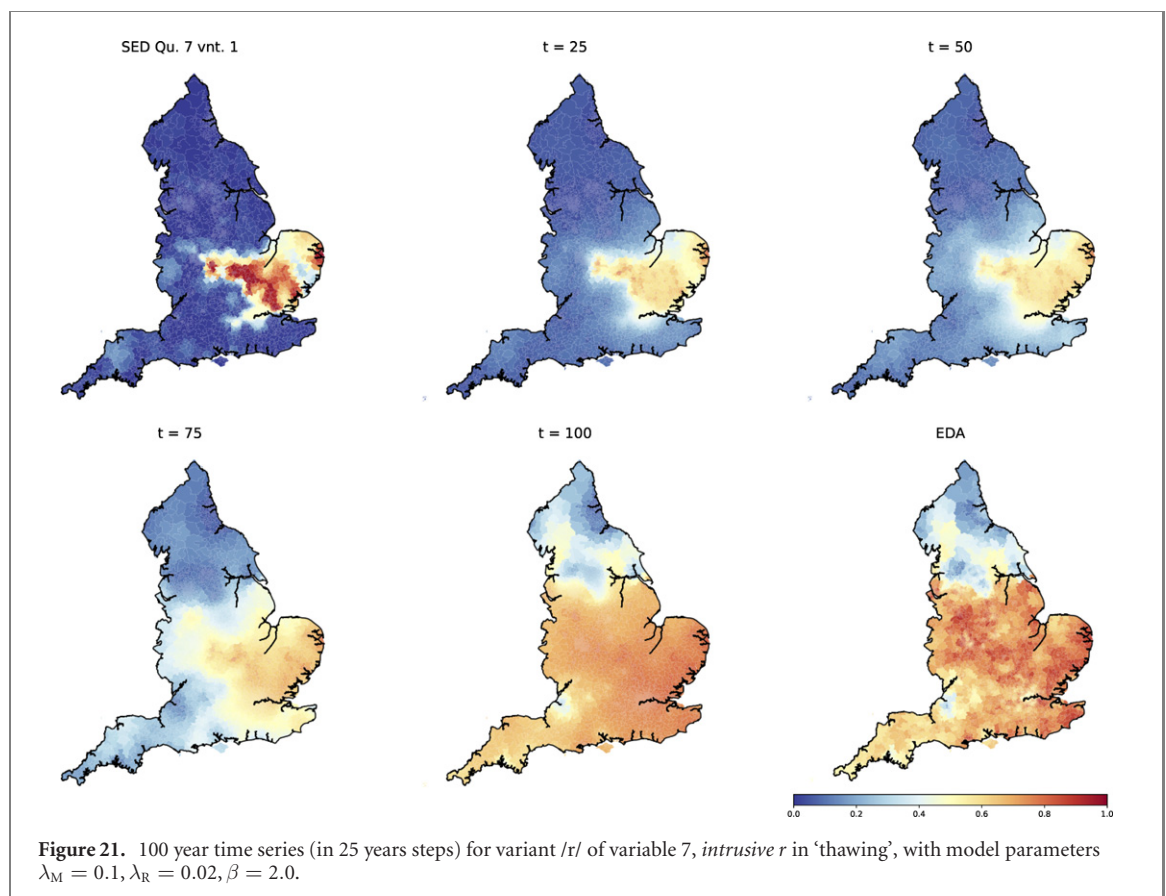
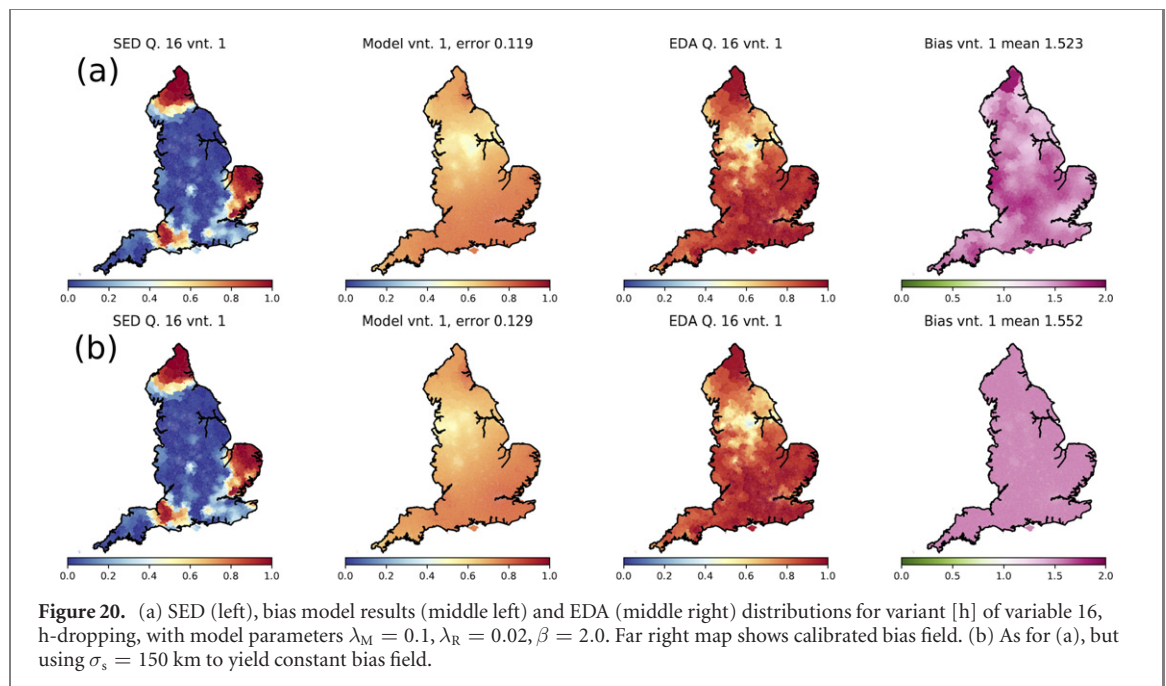


correctly predicts that the rhotic–non-rhotic isogloss in the south is retreating westwards and that rhoticity is becoming less categorical within the rhotic regions, but greatly underestimates this trend; bias in favour of non-rhoticity is required to explain how advanced the change is (see figure 19). Another example is the levelling of the 3sg.f. possessive pronoun (variable 15). The no-bias model correctly predicts that the traditional southern and Midlands variant *hern* is levelled in favour of standard *hers* to the extent that no clear isogloss is left, but incorrectly predicts remaining hotspots of traditional usage in high population density areas (Birmingham, Portsmouth/Southampton, Bristol). Bias towards the standard variant accelerates the change in the model, so that these hotspots too are levelled out. A more striking example is variable 8 (the word for *autumn*) where the no bias model predicts that the *autumn-backend* isogloss will move  $\approx 100$  km north, whereas in reality *backend* has all but disappeared, being used by only  $\approx 20\%$  of the population in the far north.

#### 5.3.4. What does bias represent?

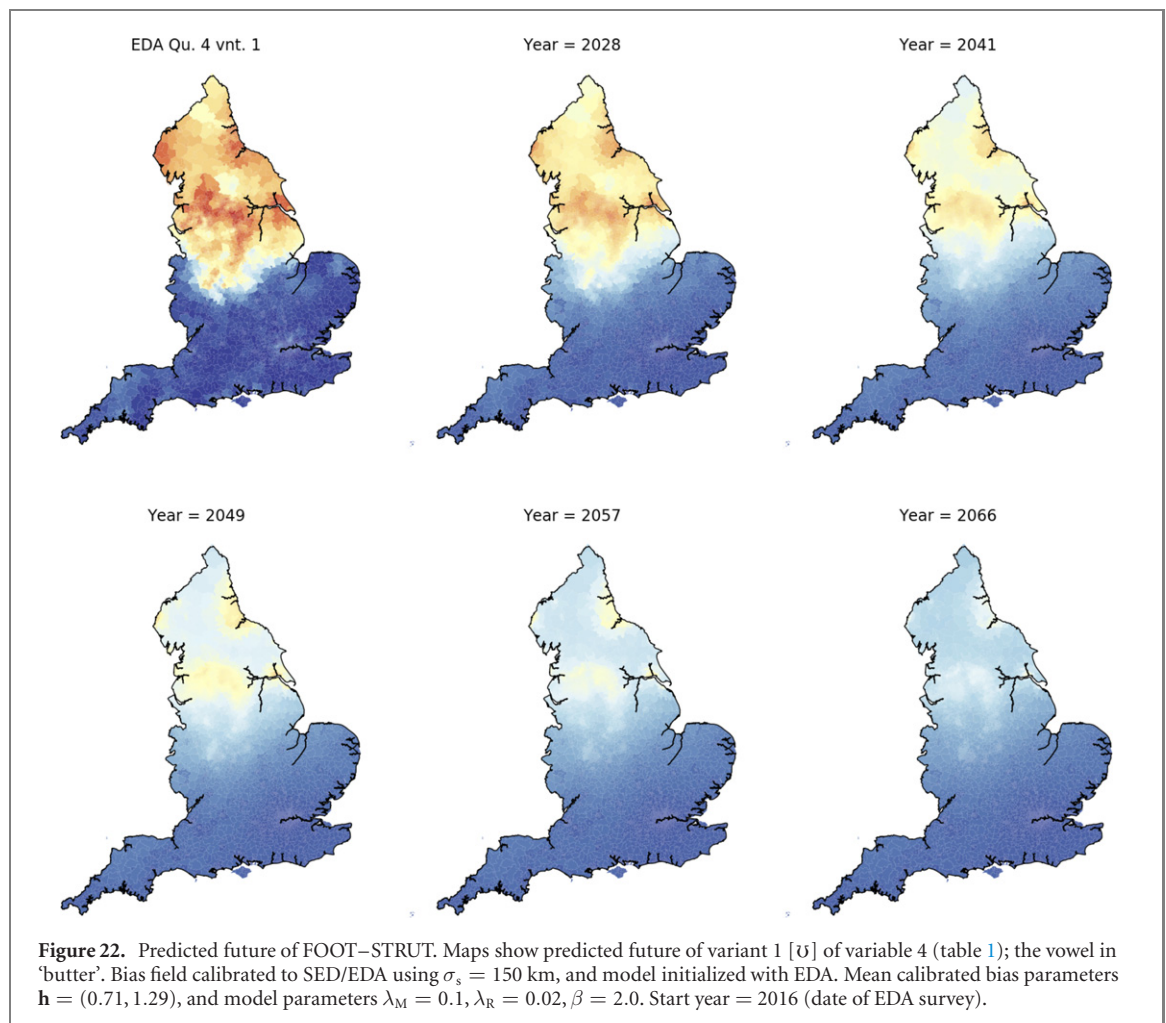
In some cases bias plausibly reflects the existence of a phonetically natural direction of change (the pronunciation of ‘l’ in shelf shifting from clear, through dark, to vocalised [ɫ], for example, but not the reverse). In principle, interaction between different dialect features could even result in such internal bias being spatially variable, although there are no absolutely convincing examples of this in these results. In rather more cases, however, it seems most likely that bias represents external (non-linguistic) factors and in particular the standard ideology. We can see this wherever bias favours a long-standing standard variant. Examples include the 3sg.m. reflexive pronoun (variable 14), rhoticity (variable 5), and the 3sg.f. possessive pronoun (variable 15). Another example is the decrease in ‘h dropping’ (variable 16) between the SED and EDA (see figure 20); because





this is a phonologically unnatural change which undoes a simplification of the pronunciation system, the bias in the model in favour of [h] may reflect the normative pressure of the standard ([h] has gained the status of *standard variant*). These results are inconsistent with an account which centres on mobility and convergence as necessary and sufficient explanatory factors for levelling. Instead, other, variant-specific factors—primarily external factors such as the standard ideology—are required.

A final pattern worth discussing is that of *innovation diffusion*. In four variables, we can plausibly label the changes we see between the SED and EDA as the expansion of a recent innovation: the [f] variant of /θ/ (variable 6); the [ʔ] variant of coda /t/ (variable 22); the [ɪ] and [ʊ] variants of coda /l/



(variable 9); and intrusive r (variable 7). Any account of the dynamics of language change—including non-spatial accounts—must offer an explanation for how innovations take hold. An innovation, by definition, begins as an extreme minority variant, and so we are faced with a challenge to explain why such variants are not immediately levelled in favour of the overwhelming majority. Possible explanations include internal linguistic reasons why particular variants might be favoured (although such explanations must also answer the question of why the change had not already taken place), and social explanations such as age vectors and incrementation or divergence for the purpose of group identity formation [21, 100, 108]. This is a larger question and not one that this paper seeks to answer; in our models, this too is simply subsumed under the bias term.

#### 5.4. Time series

So far we have presented only the initial and final states generated by our model, but we can also provide a complete reconstruction of the history of each variant over the 20th century. For example, figure 21 shows our predictions for the evolution of *intrusive r* in ‘thawing’. In principle such time series can be compared to mid-century data to validate our reconstruction of the historical changes which have taken place. Such reconstructions may have utility in historical linguistics, which aims to understand the sequences of linguistic changes which have led to the current state of the world’s languages. If we were to make an initial guess at a historical language state, a model similar to ours, adapted to allow for sequential changes in variants, could be calibrated to later states to provide a predicted history. This could then be compared to predictions made by more traditional linguistic methods. In principle it would be possible to impose additional constraints on the model, set by further data, at intervening points in its evolution. Where changes are still in progress, the model might also be used to predict the future. As an example, we have calibrated a constant bias field for variable 4 (foot–strut), giving model error 0.106. We then used this field to evolve the model into the future, using the EDA as initial condition. The results are shown in figure 22, up to the year 2066. Constant bias was used to reduce the chance of over fitting.

## 6. Conclusion

We have developed an explicit spatial model of language evolution, accounting for mundane daily mobility, realistic migration patterns, spatial distribution of population, and social conformity. One motivation is that many scholars examining diffusion and levelling in English over the last century have argued for the primacy of movement (both ‘routine’ mobility and migration) in driving change (cf section 2.3). This—tacitly or explicitly—implies that ideological and normative factors and internal linguistic pressures play only secondary roles: that widespread levelling would have taken place even without the strong standard ideology that characterises English language attitudes, and that this ideological context served only to hasten or to determine the precise direction of processes which were the product of more fundamental social-technological changes. In scientific terms, we can think of the extreme form of this view, where all change is driven by purely spatial processes (diffusion, long range mixing, interface motion), as a *null model*—the variant symmetric form of our model. We have shown that although the evolution of some variants can be explained in this way, the majority of historical changes could not have occurred without certain variants gaining special status. In some cases this bias merely accelerates or further drives changes which would have partially occurred by spatial effects alone, requiring a relatively low level of asymmetry. In other cases, strong biases are required either to drive minority variants such as *him(self)* to national dominance, or to preserve distinctive local forms such as *spelk*. In most cases the origins of these biases appear more likely to be external social factors such as normative standards, rather than internal linguistic factors. We have also seen that, even when bias is needed to explain observations, spatial effects often play an important role. Our conclusion is therefore that spatial process are important, but insufficient to explain the majority of the changes we have seen. A complete catalogue of results for all variants, including bias fields may be found in the supplementary material [104].

Beyond assessing the importance of spatial processes to language change, this modelling exercise represents a step towards the construction of minimalist spatial models which can be used to reconstruct more general historical linguistic changes, and potentially predict future change. In providing a complete hypothetical time series for a long period of time, we are offering explicit predictions which can be compared to data collected in England at any point through the century. Since the bias field is not an observable quantity, it is only these intermediate states that are directly falsifiable. We hope that such comparisons may be used to test, improve or refute our model in future.

## Acknowledgments

The authors thank the referees for their diligent and insightful reviews, which raised a number of important questions. They are also grateful for the support of a Royal Society APEX award APX\R1\180117 (funded by the Leverhulme Trust).

## Code and data availability

No new data were created or analysed in this study. Burridge is happy to be contacted with reasonable requests for additional maps, or source code.

## ORCID iDs

James Burridge  <https://orcid.org/0000-0001-6067-4187>

Tamsin Blaxter  <https://orcid.org/0000-0002-1466-8306>

## References

- [1] Castellano C, Fortunato S and Loreto V 2009 Statistical physics of social dynamics *Rev. Mod. Phys.* **81** 591
- [2] Bialek W, Cavagna A, Giardinà I, Mora T, Silvestri E, Viale M and Walczak A M 2012 Statistical mechanics for natural flocks of birds *Proc. Natl Acad. Sci.* **109** 4786–91
- [3] Yakovenko V and Dragulescu A 2000 Statistical mechanics of money *Eur. Phys. J. B* **17** 723–9
- [4] Barthélemy M 2019 The statistical physics of cities *Nat. Rev. Phys.* **1** 406–15
- [5] Baxter G J, Blythe R A, Croft W and McKane A J 2006 Utterance selection model of language change *Phys. Rev. E* **73** 046118
- [6] Burridge J and Kenney S 2016 Birdsong dialect patterns explained using magnetic domains *Phys. Rev. E* **93** 062402
- [7] Burridge J 2017 Spatial evolution of human dialects *Phys. Rev. X* **7** 031008
- [8] Derman E 2007 *My Life as a Quant P: Reflections on Physics and Finance* (New York: Wiley)

- [9] Hastie T, Tibshirani R and Friedman J 2009 *The Elements of Statistical Learning* (New York: Springer)
- [10] Blythe R A 2012 Neutral evolution: a null model for language dynamics *Adv. Complex Syst.* **15** 1150015
- [11] Nowak M A, Komarova N L and Niyogi P 2001 Evolution of universal grammar *Nature* **291** 114–8
- [12] Mitchener W G 2011 A mathematical model of prediction-driven instability: how social structure can drive language change *J. Logic Lang. Inf.* **20** 385–96
- [13] Mitchener W G 2017 A stochastic model of language change through social structure and prediction-driven instability *SIAM J. Appl. Math.* **77** 2272–93
- [14] Blythe R A and McKane A J 2007 Stochastic models of evolution in genetics, ecology and linguistics *J. Stat. Mech.* **P07018**
- [15] Blythe R A and Croft W 2012 S-curves and the mechanisms of propagation in language change *Language* **88** 269–304
- [16] Ghanbarnejad F, Gerlach M, Miotto J M and Altmann E G 2014 Extracting information from s-curves of language change *J. R. Soc. Interface* **11** 20141044
- [17] Lieberman E, Michel J-B, Jackson J, Tang T and Nowak M A 2007 Quantifying the evolutionary dynamics of language *Nature* **449** 713–6
- [18] Chambers J K and Peter T 1998 *Dialectology* (Cambridge: Cambridge University Press)
- [19] Karjy A, Blythe R A, Kirby S and Smith K 2020 Challenges in detecting evolutionary forces in language change using diachronic corpora *Glossa* **5** 45
- [20] Newberry M G, Ahern C A, Clark R and Plotkin J B 2017 Detecting evolutionary forces in language change *Nature* **551** 223–6
- [21] Stadler K, Blythe R A, Smith K and Kirby S 2016 Momentum in language change *Lang. Dyn. Change* **6** 171–98
- [22] Labov W 2001 *Principles of Linguistic Change* (Oxford: Blackwell)
- [23] Baxter G and Croft W 2016 Modeling language change across the lifespan: individual trajectories in community change *Lang. Var. Change* **28** 129–73
- [24] Kauhanen H, Gopal D, Galla T and Bermúdez-Otero R 2021 Geospatial distributions reflect temperatures of linguistic features *Sci. Adv.* **7** eabe6540
- [25] Stanford J N and Kenny L A 2013 Revisiting transmission and diffusion: an agent-based model of vowel chain shifts across large communities *Lang. Var. Change* **25** 119–53
- [26] Nerbonne J 2010 Measuring the diffusion of linguistic change *Phil. Trans. R. Soc. B* **365** 3821–8
- [27] Tesileanu T and Meyer-Ortmanns H 2006 Competition of languages and their hamming distance *Int. J. Mod. Phys. C* **17** 259
- [28] Baronchelli A, Dall'Asta L, Barrat A and Loreto V 2006 Topology-induced coarsening in language games *Phys. Rev. E* **73** 015102
- [29] Burridge J 2018 Unifying models of dialect spread and extinction using surface tension dynamics *R. Soc. Open Sci.* **5** 171446
- [30] Burridge J, Vaux B, Gnacik M and Grudeva Y 2019 Statistical physics of language maps in the USA *Phys. Rev. E* **99** 032305
- [31] Burridge J and Blaxter T 2020 Using spatial patterns of English folk speech to infer the universality class of linguistic copying *Phys. Rev. Res.* **2** 043053
- [32] Orton H and Dieth E 1962 *Survey of English Dialects* ed E J Arnold (Leeds)
- [33] Ellis A 1889 *On Early English Pronunciation with Especial Reference to Shakespeare and Chaucer Part V* (London: Trübner)
- [34] Heeringa W and Nerbonne J 2001 Dialect areas and dialect continua *Lang. Var. Change* **13** 375–400
- [35] Leemann A, Kolly M-J, Purves R, Britain D and Glaser E 2016 Crowdsourcing language change with smartphone applications *PLoS ONE* **11** 1–25
- [36] Adrian L, Kolly M-J and Britain D 2018 The English Dialects App: The Creation of a Crowdsourced Dialect Corpus *Ampersand* **5** 1–17
- [37] Vaux B and Johndal M L 2017 *Cambridge Survey of World Englishes* <http://survey.johndal.com/results/>
- [38] Grieve J, Speelman D and Geeraerts D 2011 A statistical method for the identification and aggregation of regional linguistic variation *Lang. Var. Change* **23** 193–221
- [39] Huang Y, Guo D, Kasakoff A and Grieve J 2016 Understanding US regional linguistic variation with twitter data analysis *Comput. Environ. Urban Syst.* **59** 244–55
- [40] Bloomfield L 1933 *Language* (New York: Holt, Rinehart and Winston)
- [41] Orton H and Dieth E 1962 *Survey of English Dialects* (Leeds: University of Leeds) by E J Arnold
- [42] Mathussek A 2016 On the problem of field worker isoglosses *The Future of Dialects* ed M-H Côté, R Knooihuizen and J Nerbonne (Berlin: Language Science Press) pp 99–116
- [43] Payne K J 2017 The survey of English dialects notebooks *Tradition Today: The Journal of the Centre for English Traditional Heritage* **6** 45–68
- [44] Ringe D A and Eska J F 2013 *Historical Linguistics: Toward a Twenty-First Century Reintegration* (Cambridge: Cambridge University Press)
- [45] Labov W 1994 *Principles of Linguistic Change Volume 1: Internal Factors of Language in Society* (Oxford: Blackwell)
- [46] Peter T, Gordon E, Lewis G and MacLagan M 2000 Determinism in new-dialect formation and the genesis of New Zealand English *J. Linguist.* **36** 299–318
- [47] Coates R 2018 Steps towards characterizing Bristolian *Southern English Varieties Then and Now* ed L Wright (Berlin: de Gruyter & Co) pp 188–226
- [48] Britain D 2009 One foot in the grave? Dialect death, dialect contact, and dialect birth in England *Int. J. Sociol. Lang.* **196/197** 121–55
- [49] Britain D 2002 Phoenix from the ashes?: The death, contact, and birth of dialects in England *Essex Research Reports in Linguistics* **41** 42–73
- [50] Torgersen E and Kerswill P 2004 Internal and external motivation in phonetic change: dialect levelling outcomes for an English vowel shift *J. Sociolinguist.* **8** 23–53
- [51] Tagliamonte S A 2017 Changing places: tracking innovation and obsolescence across generations *Language and a Sense of Place: Studies in Language and Region* ed C Montgomery and E Moore (Cambridge: Cambridge University Press) pp 15–37
- [52] Kerswill P 2002 Koineization and accommodation *The Handbook of Language Variation and Change (Blackwell Handbooks in Linguistics)* ed J K Chambers, P Trudgill and N Schilling-Estes (Oxford: Blackwell) pp 669–702
- [53] Kerswill P 2002 Models of linguistic change and diffusion: new evidence from dialect levelling in British English *Reading Working Papers in Linguistics* **6** pp 187–216
- [54] Kerswill P 2003 Dialect levelling and geographical diffusion in British English *Social Dialectology: In Honour of Peter Trudgill* ed D Britain and C Jenny (New York: Benjamin) pp 223–43



- [55] Britain D 2009 Supralocal regional dialect levelling *Language and Identities* ed C Llamas and D J L Watt (Edinburgh: Edinburgh University Press) pp 194–204
- [56] Auer P 2005 Europe's sociolinguistic unity, or: a typology of European dialect/standard constellations *Perspectives on Variation: Sociolinguistic, Historical, Comparative* vol 7 Trends in Linguistics 163 Berlin: Mouton de Gruyter pp 7–42
- [57] Røyneland U 2010 Vertical convergence of linguistic varieties in a language space *Language and Space: An International Handbook of Linguistic Variation, Number Bd. 30.1-Bd.30.3 in Handbücher Zur Sprach- Und Kommunikationswissenschaft = Handbooks of Linguistics and Communication Science* ed P Auer and A Lameli (Berlin: de Gruyter & Co) pp 259–74
- [58] Taeldeman J 2010 Linguistic stability in a language space *Language and Space: An International Handbook of Linguistic Variation, Number Bd. 30.1-Bd.30.3 in Handbücher Zur Sprach- Und Kommunikationswissenschaft = Handbooks of Linguistics and Communication Science* ed P Auer and A Lameli (Berlin: de Gruyter & Co) pp 355–74
- [59] Peter T 1986 *Dialects in Contact. Number 10 in Language in Society* (Oxford: Blackwell)
- [60] Giles H and Ogay T 2007 Communication accommodation theory *Explaining Communication: Contemporary Theories and Exemplars* ed B B Whaley and W Samter (Mahwah: New Jersey)
- [61] Babel M 2010 Dialect divergence and convergence in New Zealand English *Lang. Soc.* **39** 437–56
- [62] Doise W, Sinclair A and Bourhis R Y 1976 Evaluation of accent convergence and divergence in cooperative and competitive intergroup situations *Br. J. Soc. Clin. Psychol.* **15** 247–52
- [63] Pardo J S, Gibbons R, Suppes A and Krauss R M 2012 Phonetic convergence in college roommates *J. Phonetics* **40** 190–7
- [64] Peter T 2011 *Sociolinguistic Typology: Social Determinants of Linguistic Complexity* (Oxford: Oxford University Press)
- [65] Milroy L 2002 Introduction: mobility, contact and language change—working with contemporary speech communities *J. Sociolinguist.* **6** 3–15
- [66] Britain D 2013 The role of mundane mobility and contact in dialect death and dialect birth *English as a Contact Language (Studies in English Language)* ed D Schreier and M Hundt (Cambridge: Cambridge University Press) pp 165–81
- [67] Hinsken F, Auer P and Kerswill P 2005 The study of dialect convergence and divergence: conceptual and methodological considerations *Dialect Change: Convergence and Divergence in European Languages* ed P Auer, F Hinsken and P Kerswill (Cambridge: Cambridge University Press) pp 1–48
- [68] Mufwene S S 2001 *The Ecology of Language Evolution (Cambridge Approaches to Language Contact)* (Cambridge: Cambridge University Press)
- [69] Chambers J K 1992 Dialect acquisition *Language* **68** 673–705
- [70] Stuart-Smith J, Pryce G, Timmins C and Gunter B 2013 Television can also be a factor in language change: evidence from an urban dialect *Language* **89** 501–36
- [71] Watt D J L 2000 Phonetic parallels between the close-mid vowels of Tyneside English: are they internally or externally motivated? *Lang. Var. Change* **12** 69–101
- [72] Pedersen I L 2005 Processes of standardisation in Scandinavia *Dialect Change: Convergence and Divergence in European Languages* ed P Auer, F Hinsken and P Kerswill (Cambridge: Cambridge University Press) pp 171–95
- [73] Kerswill P and Williams A 2002 Salience as an explanatory factor in language change: evidence from dialect levelling in urban England *Language Change: The Interplay of Internal, External and Extra-linguistic Factors* ed M C Jones and E Esch (Berlin: de Gruyter & Co) pp 81–110
- [74] Foulkes P and Docherty G 1999 *Urban Voices: Accent Studies in the British Isles* (London: Edward Arnold)
- [75] Haug Hilton N 2010 Regional dialect levelling and language standards: changes in the Hønefoss dialect *PhD Thesis* University of York, York
- [76] Tranter M, Robineau D and Goodman G 2011 Census: population and household estimates for small areas in England and Wales *Technical Report* Office for National Statistics
- [77] Department for Transport 2014 National travel survey 2014: travel to school *Technical Report*
- [78] Paul B 2012 Education: historical statistics *Technical Report* House of Common Library
- [79] Office for National Statistics 2005 *Focus on People and Migration* ed F Ambrose, R Chappell, J Jefferies and G Wainman (London: Palgrave Macmillan)
- [80] Panayi P 2010 *An Immigration History of Britain* (London: Pearson Longman)
- [81] Bernard A 2017 Cohort measures of internal migration: understanding long-term trends *Demography* **54** 2201–21
- [82] Friedlander D and Roshier R J 1966 A study of internal migration in England and Wales: part I *Population Studies* **19** 239–79
- [83] Stillwell J and Thomas M 2016 How far do internal migrants really move? Demonstrating a new method for the estimation of intra-zonal distance *Reg. Stud. Reg. Sci.* **3** 28–47
- [84] Stillwell J C H 1978 Interzonal migration: some historical tests of spatial-interaction models *Environ. Plan A* **10** 1187–200
- [85] Wilson A G 1971 A family of spatial interaction models, and associated developments *Environ. Plan A* **3** 1–32
- [86] Batty M 1976 *Urban Modelling: Algorithms, Calibrations, Predictions* (Cambridge: Cambridge University Press)
- [87] Friedlander D and Roshier R J 1966 A study of internal migration in England and Wales *Population Studies* **20** 45–59
- [88] Trudgill P 1974 Linguistic change and diffusion: description and explanation in sociolinguistic dialect geography *Lang. Soc.* **3** 215–46
- [89] Schmidt J 1872 *The Relations of the Indoeuropean Languages* (Paris: Hermann)
- [90] Kimura M and Maruyama 1971 Pattern of neutral polymorphism in a geographically structured population *Genet. Res.* **18** 121–31
- [91] Cox J T and Durrett R 2002 The stepping stone model: new formulas expose old myths *Ann. Appl. Probab.* **12** 1348–77
- [92] Kauhanen H 2017 Neutral change *J. Linguist.* **53** 327–58
- [93] Labov W 2001 *Principles of Linguistic Change (Social Factors of Language in Society* vol 2) (Oxford: Blackwell)
- [94] Office for National Statistics 2015 *How has life expectancy changed over time?* (accessed 28 March 2021)
- [95] Feller W 1968 *An Introduction to Probability Theory and its Applications* vol 1 (New York: Wiley)
- [96] Kimura M 1953 Stepping stone model of population *Annual Report National Institute of Genetics Japan* **3** 62
- [97] Etheridge A 2009 *Some Mathematical Models from Population Genetics* (London: Springer)
- [98] Oksendal B 2010 *Stochastic Differential Equations* (Berlin: Springer)
- [99] Wright S 1943 Isolation by distance *Genetics* **28** 114–38
- [100] Michaud J 2019 Dynamic preferences and self-actuation of changes in language dynamics *Lang. Dyn. Change* **9** 61–103
- [101] Bray A J 1994 Theory of phase-ordering kinetics *Adv. Phys.* **43** 357–459
- [102] International Phonetic Association 1999 *Handbook of the International Phonetic Association: A Guide to the Use of the International Phonetic Alphabet* (Cambridge: Cambridge University Press)



- [103] MacKay D J C 2003 *Information Theory, Inference and Learning Algorithms* (Cambridge: Cambridge University Press)
- [104] Burridge J and Blaxter T 2021 Supplemental Material (<https://stacks.iop.org/JPCOMPLEX/2/035018/mmedia>)
- [105] Csiszar I and Shields P 2004 *Information Theory and Statistics: A Tutorial* (Boston, MA: Now)
- [106] Barros K, Krapivsky P L and Redner S 2009 Freezing into stripe states in two-dimensional ferromagnets and crossing probabilities in critical percolation *Phys. Rev. E* **80** 040101
- [107] Wells J C 1982 *Accents of English* (Cambridge: Cambridge University Press)
- [108] Kauhanen H 2020 Replicator-mutator dynamics of linguistic convergence and divergence *R. Soc. Open Sci.* **7** 201682